Contract No.: 500-94-0062
MPR Reference No.: 8246-080

Preliminary Report: The Impact of Prospective Payment on Medicare Home Health Quality of Care

Final Report

January 30, 1998

Arnold Chen
Helaine Noveck

Submitted to:

Health Care Financing Administration
7500 Security Boulevard (C-3-21-06)
Baltimore, MD 21244-1850

Project Officer:
Ann Meadow

Submitted by:

Division of General Internal Medicine and Geriatrics
University of Medicine and Dentistry of New Jersey/Robert Wood Johnson Medical School
125 Paterson Street, 2nd Floor
New Brunswick, NJ 08903-0019
(732) 235-7122

and

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
(609) 799-3535

Project Director:
Valerie Cheh

Principal Investigator:
Randall S. Brown

ACKNOWLEDGMENTS

We are grateful to our colleagues at MPR. Barbara Phillips and Randy Brown provided careful readings and thoughtful comments on previous drafts of this report, as well as invaluable guidance in the analysis and statistical methodology. Valerie Cheh and Jennifer Schore kindly provided much of the background material for this report. Amy Zambrowski and Rob Cedarbaum skillfully created the data files for our analyses and patiently answered our questions on the data and on computing. Walter Brower provided editorial assistance, and Jennifer Baskwell, Jill Miller, and Cindy Castro provided word processing and production support.

Dee Dee Davis of the Institute for Health, Health Care Policy, and Aging Research at Rutgers helped us with statistical computing. We thank Elizabeth Mauser Goldstein, our project officer at HCFA, for ongoing advice and assistance, and the Center for Health Policy Research at the University of Colorado for providing the Quality Assurance data and documentation.

Finally, we thank Randy Brown, Craig Thornton, and Barbara Phillips at MPR for allowing us the opportunity to work on this report. The responsibility for any errors in this report is, of course, ours.

CONTENTS

CONTENTS *(continued)*

TABLES

EXECUTIVE SUMMARY

The Health Care Financing Administration's (HCFA's) Per-Episode Home Health Prospective Payment Demonstration tests whether prospective payment for Medicare home health services increases efficiency in the provision of services while maintaining access to care and quality of care. By giving agencies the opportunity to generate profits (surpluses), prospective payment encourages agencies to reduce both the number of visits per episode and the cost per visit. Prospective payment differs substantially from current Medicare reimbursement for home health services, which pays agencies for actual costs incurred up to a specified limit and provides no rewards for efficiency.

Ninety-one agencies in five states entered the three-year demonstration at the start of their 1996 fiscal years. Prior to the start of the demonstration, the participating agencies were randomly assigned to either the treatment group (which is paid under the demonstration's prospective payment method) or a control group (which continues to be paid under Medicare's current method of cost-based reimbursement). The payments that treatment group agencies receive for the first 120 days of a patient episode are based on each agency's own costs in the fiscal year immediately preceding its entry into the demonstration, adjusted at the end of each year for changes in its case mix. During the first 120 days after admission, each agency is "at risk" for all home health visits the agency provides, but HCFA reimburses treatment agencies for up to 99 percent of fiscal-year losses. Profits in excess of specified limits must be shared with HCFA.

RESEARCH QUESTIONS AND METHODOLOGY

Though prospective payment may give agencies financial incentives to lower costs, it may also lead to undesirable effects if agencies start to "skimp" or "cut corners" in providing care. This preliminary report examines data from the first year of the demonstration to estimate impacts on quality of care. We test hypotheses on whether the demonstration has had impacts on:

- measures of patient functional health status
- patient reports of use of emergency health services
- patient admission to a hospital, a skilled nursing facility (SNF), or another home health care agency for problems potentially related to the condition under the care of the demonstration agency
- patient mortality

Quality Assurance (QA) Data

We studied two separate analytic datasets of episodes of home health care, the QA data, and the Claims data. The QA data contains information on patient functional health status at two points in time and on reported use of emergency health services. These data were collected by the demonstration QA contractor, the Center for Health Policy Research at the University of Colorado. Four of the 91 demonstration agencies were so delayed in implementing QA data collection that they had insufficient data for this analysis. Furthermore, to avoid disproportionate influence on the results from very small agencies, we excluded five with fewer than 20 episodes each, leaving 82 agencies. The analysis of QA data reported here is based on roughly 29,000 episodes of home health care with dates of admission to home care ranging from early May 1996 to late April 1997.

The interpretation of treatment-control comparison is difficult because of two sources of systematic bias in how outcomes are measured in the QA data. The outcomes in the QA data consist either of changes in patient health status or of events that occur between home health admission and a follow-up data collection point. For example, patients are classified as having Improved (yes/no), or Stabilized (yes/no) between admission and followup in 17 areas of health (for example, Pain, Pressure Ulcer Count, Ambulation/Mobility). They are noted as having reported any Emergency Health Service Use (yes/no) such as emergency visits to hospital emergency rooms or doctors' offices between admission and followup. The first source of measurement bias in the QA data occurs because the timing of the followup data collection depends on the length of the episode, occurring at discharge or 120 days after admission, whichever occurs first. Because treatment agencies discharge their patients more quickly than control agencies (Cheh et al. 1997), the follow-up data collection point for treatment agency patients is earlier in the course of home health care than for control agencies. Depending on the underlying trajectory of recovery of the health item being measured, even identical outcomes of treatment agency patients may *appear* to be different simply because they are measured at different time point on the trajectory.

The second source of measurement bias in the QA data stems from differences between treatment and control agencies in visit frequency. Visit frequency may influence agency staff ascertainment of such outcomes as frequency of behavioral problems or emergency health care use simply because increased contact with and prompting by agency staff can affect patient/family recall of such events. Treatment agencies provide fewer home visits, and thus apparent demonstration effects can arise without any true underlying impacts.

We analyzed the QA data with unadjusted comparisons of treatment and control group means. The home health agencies varied widely in size, but we weighted observations to give each agency equal representation in the analysis. Standard errors were calculated using special software that accounts for the effects of sample clustering and weighting, so as to avoid overstating the precision of the estimates. We examined the sensitivity of the results to the alternative weighting scheme of representing each agency proportional to its size.

Claims Data

The second, or Claims, dataset contains information on admission during the risk period to a hospital, a SNF, or another home health care agency for a diagnosis involving the same body system as that being cared for by the demonstration agency. For example, a patient receiving home health care for a disease of the circulatory system admitted during the risk period to a hospital, a SNF, or another home health care agency for a circulatory system problem is considered to have an adverse outcome, but an admission for a disease of the genitourinary system would not count as an adverse outcome. Such admissions for the same general problem being treated in home health care are more likely to reflect quality of care problems in the home care agency. The Claims data also contain mortality outcomes. Data on the outcome variables were drawn from Medicare claims files. Data collected at patient admission for case-mix adjustment and from earlier Medicare claims provided measures of preadmission characteristics of patients admitted to demonstration agencies. Data on agency characteristics were obtained from the agency cost reports and the demonstration implementation contractor.

The Claims data analysis reported here is based on roughly 51,000 home health episodes taking place in 87 of the demonstration agencies (4 of the 91 agencies were excluded because they dropped out of the demonstration.) All admissions occurring between an agency's demonstration start date and August 1996 are included.

We used multiple logistic regression to model the Claims data outcomes as a function of agency treatment or control status while controlling for baseline differences between patients, agencies, and health care markets. This approach proved crucial to obtaining valid impact estimates since, despite the randomization of participating agencies, there were several significant differences between treatment and control agencies aside from the method of payment. As in the QA data analysis, observations were weighted so that each agency was represented equally in the analysis and standard errors of impact estimates inflated for the effects of sample clustering and weighting. The robustness of results was also checked with the alternative weighting scheme of agencies represented in the analysis proportional to size. The institutional admission and mortality outcomes of the Claims data may lack the richness of the QA data's functional and health status outcomes, but a powerful advantage of the Claims data is that outcomes are measured comparably between treatment and control agencies.

FINDINGS

QA Data

We found few significant differences in patient outcomes between treatment and control agencies. In the QA data, we performed unadjusted comparisons of 17 measures of Improvement in health or function (for example, Improved in Pain, Improved in Ambulation). Only two of the 17 comparisons were statistically significant: Behavior Problem Frequency and Urinary Incontinence or Catheter Present, both favoring the treatment group. Of the 17 corresponding Stabilization rates (for example, Stabilized in Pain, Stabilized in Ambulation), small but statistically significant differences in Behavior Problem Frequency (favoring treatment agencies) and

Transferring (favoring control agencies) were found. Finally, there was a statistically significant lower rate of hospital emergency room visits for the treatment agencies. Because of the measurement biases described above, treatment-control differences in the QA data are difficult to interpret, but at least those that were found were of small magnitude. The results were not substantially affected when reanalyzed using weights that represented agencies weighted proportional to size.

Claims Data

Results from analysis of the Medicare claims data suggest no adverse effects of the demonstration on the quality of care. We found treatment group patients to have significantly *lower* rates of admission to emergency care than control agency patients; and found no significant differences between treatment and control agencies for any of the other outcomes, regardless of whether observations were weighted to represent agencies equally or in proportion to their size. Although the measurement biases make the QA data difficult to interpret in isolation, the lack of program impacts in the independent and unbiased Claims data, as well as the separate findings by Schore (1997), corroborate the conclusion that the demonstration has had no impact on any of the outcomes measured.

IMPLICATIONS OF THE FINDINGS

We find no evidence that quality of care as measured by patient outcomes has suffered as a result of the demonstration. The separate reports by Cheh et al., "Preliminary Report: The Impact of Prospective Payment on Medicare Home Health Use: Promising Results for a Future Program" (1997), and Schore, "The Impact of Prospective Payment on Medicare Service Use and Reimbursement During the First Demonstration Year" (1997), show the demonstration has substantially reduced home visits without increasing use of other Medicare services. Our report and those of Cheh and Schore, all of which are based on only early demonstration data, lead to the conclusion that per-episode prospective payment for home health care could lower Medicare's costs of providing home care services without harming quality of care.

This interim report is subject to limitations. The findings are based only on early demonstration data, and impacts on patient outcomes from changes in agency behavior might not become discernible until later in the demonstration. In addition, our statistical power to detect small impacts is limited. Caution should be exercised in generalizing these results to home health agencies not participating in the demonstration or to a national program of per-episode payment for Medicare home health care. As in any study in which the participants are volunteers, demonstration agencies may be those best able to respond to the incentives of the demonstration (for example, they may be more concerned than others with the quality of care they deliver or better able to reduce visits without adversely affecting patients). The evaluation's final report will include an analysis of the representativeness of participating agencies.

We also note that the results may not provide a reliable guide to how agencies will behave and their patients fare under a national program of per-episode payment, because such a program would likely differ from the demonstration. Under a national program, agencies would not be protected

from incurring financial losses, which could compel some to respond more aggressively. Furthermore, the per-episode rate paid to an agency would probably not be based on its own prior average, greatly increasing the potential for losses for agencies that tend to provide large numbers of visits per episode or to have high costs. This might cause agencies to reduce visits more than we observed in our preliminary analysis and, in turn, may lead to lower quality of care. On the other hand, such pressures may be offset by the commitment of agencies to their patients, by pressure to keep patients and referral sources satisfied, and by HCFA's quality review process.

Despite these concerns, our results and those of Cheh et al. (1997) and Schore (1997) suggest that agencies have responded to the financial incentives of the demonstration in a way that reduces home health visits and episode length without adversely affecting patients. The findings are encouraging for policymakers seeking to improve efficiency in Medicare home health care while maintaining the quality of care.

Further evidence will come from the final evaluation report on quality of care. That report will include data from all three demonstration years and will examine differences in impacts across important subgroups. A major focus of the final report will be the analysis of the patient survey data. Randomly selected home care patients in the demonstration have been surveyed by phone three months and eight months after home health admission on health status and function, satisfaction with home health agency care, and agency quality of care. A crucial advantage of the survey data is that they are collected at a uniform point in time after home health admission and are thus free of the measurement biases of the QA data.

I. THE PER-EPISODE HOME HEALTH DEMONSTRATION AND EVALUATION

HCFA's Per-Episode Home Health Prospective Payment Demonstration tests the extent to which prospective payment for Medicare home health services increases efficiency in the provision of services, reduces public expenditures, and maintains access to care and quality of care. Per-episode payment encourages participating agencies to generate savings by reducing the number of visits per episode and the cost per visit. These incentives differ greatly from those of the current system of cost-based reimbursement, which provides no reward for efficient care delivery. However, reductions in visits per episode and in cost per visit could have adverse effects on the quality of patient care.

The demonstration encourages home health agencies to increase efficiency and generate surpluses by reducing home health visits. Given the federal mandate to introduce prospective payment nationally for Medicare home health care, it is critical to know whether agency responses to the incentives of per-episode payment lead to worse patient outcomes. This report presents the findings of an interim analysis of demonstration impacts on quality of home health care as measured by patient outcomes and health services utilization.

The first chapter provides an overview of the history of the Medicare home health benefit and the Medicare-certified home health industry, followed by a description of the Per-Episode Home Health Prospective Payment Demonstration. The reader who is familiar with Medicare home health may wish to begin with Section B, which describes the demonstration. Section C provides an overview of the research issues and general approach taken to conducting the evaluation. Section D describes the hypothesized effects of per-episode payment on the quality of home health services.

A. THE MEDICARE HOME HEALTH BENEFIT

When Congress created the Medicare program in 1965, it included a home care benefit. Home health benefits were included as post-acute care under Part A as an alternative to expensive inpatient care. In efforts to increase access to the benefit, lawmakers have modified it a number of times since its inception.

Currently, the Medicare home health benefit covers home health services under Parts A and B. There is neither a deductible nor a coinsurance payment. To be eligible for home health benefits, the beneficiary must (1) have Medicare coverage; (2) be homebound; (3) be under the care of a physician; and (4) need skilled nursing, physical therapy, or speech therapy on a part-time or intermittent basis.[1]

HCFA administers the Medicare home health benefit through fiscal intermediaries (FIs), each serving a defined geographic region of the country. In addition to serving as communication links between HCFA and the agencies, FIs also review claims to limit inappropriate use of services, determine reasonable costs, and administer payments to home health agencies.

Outside the prospective payment demonstration, Medicare reimburses agencies for the reasonable costs incurred to provide care. During the period covered by this report, an agency's costs were judged reasonable as long as they did not exceed 112 percent of the mean cost incurred by all agencies (for the agency's mix of visits) in the same geographic area. Agencies incurring

---

[1]Skilled nursing services are covered as long as (1) a physician has ordered them, (2) they are required on a part-time or intermittent basis, (3) they require the skills of a registered nurse (or a licensed practical nurse or licensed vocational nurse under a registered nurse's supervision), and (4) they are reasonable and necessary to treat an illness or injury. Physical therapy and speech therapy are covered if recommended by a physician. Beneficiaries who need only occupational therapy are entitled to benefits only if they have established a prior need for skilled nursing care, speech therapy, or physical therapy in the current or prior certification period (see Teplitsky and Janson 1985-1992, p. VII.23, Section 204.4).

aggregate costs that exceed these limits were reimbursed only up to the limits (known as the Section 223 limits). The cost limits were frozen for reporting periods that began between July 1, 1994, and June 30, 1996.[2]

Expenditures for home health care represented a substantial nine percent of all Medicare expenditures in fiscal year 1994 and have been rapidly growing in recent years (HCFA 1997). Spending for Medicare home health services has grown at least 20 percent a year since 1989, the year in which coverage was broadened as part of the settlement of a lawsuit brought against HCFA. After a 53 percent spike in annual growth in 1990, however, the rate of growth has declined (ProPAC March 1996). Little of the growth is due to increases in cost per visit; rather, it is due to increases in the number of beneficiaries receiving Medicare home health care and the number of visits per beneficiary.

Since Medicare's inception, the number of Medicare-certified agencies has more than quadrupled. In 1995, there were roughly 8,700 Medicare-certified home care agencies (ProPAC March 1996). Administratively, home health agencies have different ownership and auspices. They can be freestanding for-profit, freestanding nonprofit, affiliated with a facility (such as a hospital or skilled nursing facility), or operated by a government entity. Most of the recent growth in the number of Medicare-certified agencies has been in the number of hospital-based and freestanding for-profit agencies (ProPAC March 1996). The distribution of ownership/auspices varies considerably by region of the country. Government-operated and private nonprofit agencies dominate the Northeast. Freestanding, for-profit agencies are pervasive in the South and West and even dominate the markets in some states.

---

[2]The Balanced Budget Act of 1997 provides for a different approach to setting the limits with cost reporting periods beginning October 1, 1997. The act also extends the freeze on the cost limits that began July 1, 1994.

The number of Medicare-covered visits per episode and the length of episodes also vary widely across regions. For example, the national averages in 1990 and 1991 for number of approved visits in an episode of home health care was 47 and for episode length 94 days, but the mean number of visits per episode varied from 28 in the Pacific region to 95 in the East South Central region, and mean episode length varied from 60 days in the Pacific region to 180 in the East South Central region (Schore 1995). In 1994, the mean visits per beneficiary served was 66 nationally but varied from 45 in the Pacific region to 106 in the East South Central region (HCFA 1996).

The dramatic growth of home health as a proportion of total Medicare spending, combined with striking regional variation in its use and the explosive growth of the home health industry, prompted Congress to legislate changes to the Medicare home health benefit as part of the Balanced Budget Act of 1997. The act mandates the implementation of per-episode prospective payment for Medicare home health by 1999.

B. THE PER-EPISODE DEMONSTRATION

During the period studied here, the Medicare home health payment system reimbursed agencies for allowable costs up to a limit of 112 percent of the mean national cost. Because there was no mechanism for home health agencies to realize savings beyond costs, this system provided no incentives to produce services efficiently and, in effect, subsidized inefficient providers. Per-episode prospective payment is meant to increase efficiency, using the opportunity to generate savings as the primary incentive.

Ninety-one Medicare-certified home health agencies in five states--California, Florida, Illinois, Massachusetts, and Texas--enrolled in the three-year per-episode demonstration.[3] Forty-seven of

[3]Reflecting the United States more generally, considerable variation existed in the use of (continued...)

them were randomly assigned to the treatment group to receive per-episode payment. The remaining 44 were assigned to the control group to continue under cost reimbursement. The first agencies in the treatment group began implementing prospective payment in June 1995; the latest entrants began in January 1996. Each agency started as its fiscal year began. Demonstration operations will continue through December 1998.

Mathematica Policy Research, Inc. (MPR) is the evaluation contractor responsible for assessing the impacts of the demonstration and its implementation. Several other organizations are participating in the demonstration. Abt Associates, Inc., is the implementation contractor responsible for recruiting demonstration agencies, monitoring the status of demonstration operations, and calculating certain statistics needed for agency payment. Palmetto Government Benefits Administrator (PGBA) is the FI responsible for review of claims and agency payment. The Center for Health Policy Research (CHPR) at the University of Colorado is the QA contractor responsible for designing and implementing a QA system for the demonstration agencies.

1. Demonstration Payment and Incentives

HCFA developed the Home Health Prospective Payment Demonstration to assess whether the profit motive can increase the efficiency of providing Medicare home health care and thereby reduce public expenditures, without sacrificing access to care or the quality of care. Phase I of the demonstration, which tested per-visit prospective rate setting, provided agencies an opportunity to

---

[3](...continued)
Medicare home health across the five demonstration states. In 1995, the mean numbers of visits provided per beneficiary using home health were as follows: California, 54; Illinois, 55; Florida, 81; Massachusetts, 94; and Texas, 117 (HCFA 1997).

generate profits (and avoid losses) by reducing per-visit costs.[4] The current phase of the demonstration, Phase II, tests per-episode prospective payment. Under per-episode payment, agencies may earn profits by reducing the number of visits, as well as by reducing per-visit costs.[5]

a. Payment

Agencies selected for the treatment group receive a lump-sum payment for the first 120 days of home health care, regardless of number of visits provided or their cost.[6] The agencies are thus "at risk" for the costs of care incurred during this period. Those agencies that can provide care for less than the fixed (per-episode) rate will generate profits, whereas those whose costs exceed the fixed rate will incur losses. However, as described below, HCFA reimburses agencies for up to 99 percent of their losses, so there is essentially no risk of losses as long as agencies hold their costs below the limits.

For each visit beyond 120 days (referred to as outlier visits), treatment agencies receive a fixed, per-visit payment rate that varies by the type of visit. In the demonstration, a treatment agency is also paid on a per-visit basis for visits made to patients admitted before the agency began demonstration operations ("phase-in" visits) and to those admitted within 120 days of the end of

---

[4]The per-visit demonstration was implemented in the same five states; however, most of the agencies participating in the per-episode demonstration did not participate in the per-visit demonstration. (Only agencies in the per-visit control group were eligible.)

[5]Strictly speaking, only for-profit agencies earn profits; nonprofit agencies generate surpluses. However, for brevity, we use the term "profits" in this report to refer to surpluses generated by nonprofit agencies as well as profits earned by for-profit agencies.

[6]Durable medical equipment, nonroutine medical supplies, and Part B home health services (for people not eligible to receive Medicare home health care under Part A) continue to be reimbursed at cost throughout the demonstration. In addition, if an agency did not provide one or more of the six Medicare services during the base year but begins to do so during the demonstration, those visits are also reimbursed at cost, as are the costs of care for which Medicare is a secondary payer.

demonstration operations in that agency ("phase-out" visits). Agencies that can provide an outlier, phase-in, or phase-out visit for less than the fixed (per-visit) rate can also generate profits.

In the demonstration, home health episodes are defined by gaps in Medicare-covered home health care of at least 45 days. Only after the 120-day risk period and a 45-day gap in services can an agency receive a new per-episode payment for a given Medicare beneficiary.

b. Rate Setting

Prospective per-episode rates are based on an agency's costs and episode profile in the fiscal year preceding its entry into the demonstration (the base year), adjusted for inflation and changes in case mix in each evaluation year. The episode profile is the average number of visits provided by the agency during the first 120 days of an episode, calculated for each of the six types of visits covered by Medicare. This episode profile is "costed out" at the agency's per-visit costs in the base visit year, adjusted for inflation. Payment for outlier, phase-in, and phase-out visits are also based on the agency's base-year per-visit costs (adjusted for inflation).[7] HCFA's market basket is used to adjust both the per-visit and per-episode rates for inflation.

The case-mix adjuster classifies each patient into one of 18 groups on the basis of 12 variables that describe the patient's characteristics. From this information, an aggregate case-mix index is created for each agency. At the end of each year of the demonstration, an agency's case-mix index for that year is compared with its case-mix index in the base quarter (the last quarter of the base year). If the agency's case mix differs, its aggregate payment is retrospectively adjusted.

[7]Because complete data for episode profiles and settled cost reports are not available for a given year until some months after that year is over, the initial lump-sum and per-visit rates used in the demonstration were preliminary and were revised as final base-year data became available.

c. Loss Sharing and Profit Sharing

To encourage agencies to participate in the demonstration, HCFA provided a loss-sharing arrangement. HCFA reimburses treatment agencies for 99 percent of losses in the first demonstration year, and for 98 and 97 percent of losses in the second and third years of the demonstration, respectively, as long as total payments do not exceed the Section 223 limits.

To counteract the incentive to reduce the quality of care to generate profits, as well as to prevent agencies from realizing windfall profits at public expense, HCFA shares in profits above a specified threshold profit rate. If the total of a treatment agency's per-episode and per-visit prospective payments is greater than the costs incurred in rendering the services covered by these payments, profit greater than five percent of total allowable costs for these services is subject to profit sharing with HCFA. HCFA's share of profits is 25 percent of profits over 5 percent but less than or equal to 15 percent of total allowable costs for these services, plus a larger share of any profits over 15 percent (with the share of profits over 15 percent varying by demonstration year).

d. Incentives

Treatment agencies can reduce the cost of care rendered during the 120-day period by (1) reducing the number of visits provided during the risk period, (2) changing the visit mix to make less-costly visits a larger proportion of the total number, or (3) reducing per-visit costs (or some combination of these three). Reductions in the number of visits during the risk period could involve discharging patients earlier, thereby reducing the length of the episode or reducing the frequency of visits without reducing episode length. Reductions in the average number of visits could also be achieved by admitting a mix of patients needing less care. Reductions in per-visit costs could be achieved by cutting either direct costs (such as the length of a visit) or administrative costs (such as supervision). Alternatively, agencies might accept increases in per-visit costs to reduce the number

of visits during the risk period. For example, agencies might hire wound care specialists (who command higher salaries) and thereby reduce the number of visits, or they might use additional administrative resources to monitor the number of visits provided, thereby increasing per-visit costs. Per-visit costs might also increase if agencies perform in a single longer visit services that they previously provided in (and billed as) two separate visits. In addition, as agencies reduce per-episode visits, they may experience a reduction in the direct-cost base over which their administrative costs must be spread, which may mean some loss in economies of scale. As a result, treatment agencies have an incentive to increase the number of outlier (and phase-in and phase-out) visits to help offset any volume reductions due to decreases in the number of visits during the risk period, as well as to increase the number of patients they serve.

Profit motive is the prime incentive offered under the demonstration. While treatment agencies may incur losses, the generous loss-sharing provisions of the demonstration limit the incentive for agencies to alter their behavior to avoid losses, particularly in the first demonstration year. Thus, the demonstration's incentives rely heavily on the "carrot" of profits and relatively little on the "stick" of losses.

Agencies' responses to the incentives offered by the demonstration will depend on the priority each agency places on maximizing profits, relative to other goals. Nonprofit agencies in particular may view their primary mission as meeting the needs of the communities they serve. Consequently, they may be more reluctant than for-profit agencies to reduce visits during the risk period on the grounds that doing so would reduce care to those in need. The demonstration, however, does provide nonprofit agencies with an opportunity to generate profits that could then be used to develop

programs of benefit to their community or to provide services to community members who cannot obtain them in other ways (such as through Medicaid, other public programs, or private purchase).[8]

Given the demonstration's emphasis on profit motive, we expect for-profit agencies to respond more aggressively than nonprofit agencies to the incentive of per-episode prospective rate setting. We also expect that hospital-based agencies may be less responsive than freestanding agencies to the opportunity to earn a profit under the demonstration. The latter must respond to the hospital's need to discharge patients promptly and to "flow down" hospital administrative costs to the home health agency. Attention to the needs of the parent organization may also affect the behavior of other agencies that belong to a chain or other system of organizations.

2. Other Demonstration Procedures

a. Medical Review

For agencies in the treatment group, the demonstration FI performs only limited medical review (known as "abbreviated" medical review) is for care delivered during the risk period. This review seeks to determine whether the patient met the coverage criteria for home health care and received at least one visit that met these criteria. Only the admission bill is reviewed. As a condition of payment, the demonstration FI requires that the agency submit HCFA 485 and 486 forms or clinical notes for admissions that coincide with an episode eligible for prospective payment. The medical review process is based on these materials.

---

[8]It is theoretically possible that nonprofit agencies might take advantage of the loss-sharing provisions to increase visits during the risk period, if they believe that Medicare has restricted the provision of needed care. Nonprofit agencies might treat the loss-sharing provisions as a source of community service funds, accessible with a small amount of funding (equal to one, two, or three percent of losses) from private sources.

All visits paid for under per-visit rate setting are subject to the usual focused medical review, under which a sample of claims is reviewed to ensure that each visit is medically reasonable and necessary. Medical review for control agencies continues under the current (nondemonstration) regulations. The only major difference is that control agencies are assigned to the demonstration FI. Since the demonstration FI's medical review procedures may differ in minor ways from those of other FIs, control agencies may be subject to policies somewhat different from those they are accustomed to.

b. Billing

Treatment agencies must submit an admission bill to the demonstration FI to initiate an episode of care. They are then expected to submit interim bills for the rest of the risk period, although the episode payment to the agency depends only on the admission bill.[9] The agency must bill separately for any outlier visits. When a patient is discharged, during either the risk or the outlier period, agencies are to submit a discharge bill to terminate the episode. The FI will not initiate a new episode for a given if a prior episode has not been terminated. In addition, the FI checks that the 120-day risk period and a 45-day gap have elapsed.

If a treatment admission claim is accepted (following abbreviated medical review), the per-episode payment is made as a lump sum.[10] While medical review is pending, subsequent episode

[9]The interim bills are required for reimbursement for supplies and proper calculation of costs for profit and loss sharing with HCFA. Interim bills also provide information on the number of visits, which is required for the evaluation.

[10]If the admission claim is denied, interim claims for that episode are suspended for 65 days to await appeal. If an appeal is filed, interim claims are suspended also until a decision is made on the appeal for the admission claim. When an admission claim is denied and an appeal is not filed within 65 days, or if the denial of the admission claim is upheld on appeal, suspended interim claims are released for possible payment under per-visit rate setting.

bills are suspended. Initially, all episodes were subject to abbreviated medical review; in mid-1996, however, the proportion was reduced to 25 percent.[11]

Periodic interim payments (PIPs), which are intended to smooth cash flow for home health agencies, were originally discontinued for treatment agencies. However, a similar periodic payment system (called biweekly interim payments [BIPs]) was later reintroduced to meet the cash flow needs of some treatment agencies whose receipt of per-episode payments was delayed.

Control agencies continue to submit bills as under cost reimbursement and continue to be eligible for PIP. The demonstration FI bases PIP payments on the average cost for each type of visit, while other FIs base PIP payment on overall average cost per visit. As a result, there may be minor differences in control agency PIP payments compared to what control agencies have experienced outside the demonstration.

c. QA

All agencies participating in the demonstration (in both the treatment and control groups) are required to collect and submit patient-specific information to the demonstration QA contractor. The QA procedures are designed to produce "continuous quality improvement." Visiting staff from demonstration agencies are required to collect information (primarily on functional status and medical condition) at admission or a return to home health care following an inpatient stay of 48 hours or more, and then again at discharge, 120 days after admission, or the last home visit before an inpatient stay of 48 hours or more, whichever comes first. The QA contractor uses this

[11]Abbreviated medical review was required for all episodes during most of the early months of the demonstration included in this analysis.

information to develop profiles describing patient outcomes at each agency. These profiles are provided to the demonstration agencies to help them improve the quality of care they provide.

C. COMPONENTS OF AND APPROACH TO THE EVALUATION

The evaluation seeks to answer two broad sets of policy questions: (1) How did home health agencies make decisions about participating in the demonstration and how did they respond to its incentives? and (2) To what extent did per-episode prospective payment for home health affect the costs of home health agencies and outcomes for their patients--that is, what was its impact? These areas of inquiry are interrelated. Impacts can be interpreted only through understanding of how participating agencies changed their practices under per-episode prospective payment and how the demonstration operated. It is critical to understand which strategies help produce more-efficient home health care and how government can best shape policies to encourage such behavior by agencies with widely differing missions and characteristics. In addition, the nature and extent of program impacts must be determined so that analysis of decisions and operations can be focused on agencies that were particularly successful (or unsuccessful) in improving efficiency. The integration of these two analyses is a key aspect of our evaluation. (For a detailed discussion, see Phillips et al. 1995.)

1. Analysis of Agency Decisions and Operations

The analysis of agency decisions and operations focuses primarily on five questions:

1. What factors explain the decision of home health agencies to participate in the demonstration?
2. Was the demonstration implemented as planned?
3. What strategies did agencies adopt to reduce visits per episode and per-visit costs?

4. What were the effects of these strategies on the process of care, and what implications do these effects have for access to care and the quality of care?

5. What factors explain the key decisions of home health agencies?

The evaluation took a case study approach to answering these questions. This approach included judgmentally selecting 67 of the agencies participating in the demonstration for site visits at the start and conclusion of the demonstration. The goal in selecting agencies for site visits was to obtain as much information as possible about various agency environments, decision-making processes, and responses to the demonstration. The site visit data are supplemented with program records and discussions with other demonstration actors.

The following were the key conclusions of the first report on agency decisions and operations (Phillips and Thompson 1997), which was based on site visits conducted during the first half of 1996 (and the supplemental material just described):

- Most agencies saw the demonstration as an opportunity to learn about operating under prospective payment with limited financial risk. Over half (including both for-profit and nonprofit agencies) saw generating a profit or surplus as another key objective.

- Most agencies characterized the home health environment in which they operated as highly competitive, although hospital-based agencies perceived competition as less intense (probably because they had "protected" referral streams). Nearly half the agencies we visited reported an oversupply of agencies in the communities they served.

- More than 80 percent of agencies had Medicare managed care plans operating in the areas they served. This varied from 100 percent of California and Florida agencies to 70 percent of Texas agencies. Even though managed care was widespread, however, most demonstration agencies served few managed care enrollees.

- Agencies expected that growth rates during the demonstration would be dramatically different from those that prevailed only a few years ago. About half expected to increase the number of patients seen, with anticipated growth rates averaging about nine percent a year. About 15 percent of agencies expected to shrink, with expected declines in caseloads averaging about 8 percent.

- Treatment and control agencies planned to reduce their costs per visit during the demonstration. The most common strategy planned was to reduce administrative costs; another approach was to increase use of technology.

- About half of treatment agencies planned to reduce per-episode costs. Strategies included reviewing utilization more intensely, rationalizing the process of care through care maps (or critical pathways), placing greater reliance on community services and family caregivers, and increasing use of telephone contact with patients.

In a separate report we will investigate the implementation of demonstration procedures and responses to demonstration incentives as they evolved during the three-year demonstration. We will also prepare a quantitative analysis of the factors associated with agency participation in the demonstration.

2. Analysis of Program Impacts

In this section, we provide an overview of the issues, data, and methods for the overall analysis of program impacts. The methods and data for the preliminary analysis of impacts on home health quality of care are described in detail in Chapter II.

a. Research Issues

Many of the critical policy issues to be addressed in the evaluation pertain to program impacts--the extent to which per-episode prospective payment for home health care affects the behavior of home health agencies and outcomes for their patients. Controlling public expenditures for home health care would be a primary objective of a national program of prospective payment for home health agencies. A key aim of the evaluation, therefore, is to measure the impact of per-episode prospective payment on per-episode service cost, to determine the potential for savings. Because per-episode prospective payment may alter per-visit costs and the mix of visits rendered, as well as the number of home health visits, it will be necessary to identify the relative importance (to any

expenditure reductions) of changes in the cost of producing a visit and in the number and types of visits rendered. It will also be important to assess whether effects on service use and agency behavior are likely to affect access to or quality of care, as well as the extent of any such effects. Through its potential effect on access to and quality of care, per-episode prospective payment may shift care to nursing homes or hospitals, to programs that provide community-based services, or to informal caregivers (that is, family members and friends). The evaluation will also identify the extent of such shifts and implications for the overall burden and cost of care borne by public programs (including Medicaid) and informal caregivers.

These policy issues suggest the following key research questions concerning demonstration impacts:

- What effect does per-episode prospective payment have on Medicare home health services received during the risk period, the outlier period, and overall?
- What effect does per-episode prospective payment have on per-visit costs for Medicare-certified home health agencies and on the volume and types of services these agencies provide?
- What effect does per-episode prospective payment have on patient selection and retention and thus on access to care?
- What effect does per-episode prospective payment have on quality of care?
- What effect does per-episode prospective payment have on Medicare expenditures generally?
- What effect does per-episode prospective payment have on the use of and expenditures for non-Medicare-covered services, including the use of Medicaid services, other home- and community-based services, and informal care?
- Do the effects of per-episode prospective payment vary with the characteristics of the patient or the agency?

b. Methods and Data

Our approaches to estimating impacts on different outcomes are similar in principle; they differ only in the unit of analysis, the samples and data sources, and the statistical model. For each outcome measure, we will compare the experience of the treatment agencies (or their patients) with that of the control agencies (or their patients) during the demonstration period.

For most of the analyses, the episode will be the unit of analysis. The fact that payment will be fixed for an episode provides a compelling reason for using the episode as the analysis unit. While patients could be used as the unit, some will have multiple episodes. We will also use the agency as the unit of analysis for investigating impacts on agency outcomes, such as cost per visit and structural aspects of quality.

Outcome measures are drawn from secondary and primary sources. The secondary sources include Medicare Standard Analytical Files (SAFs) and enrollment files, data collected by the demonstration QA contractor, State Medicaid Research Files (SMRFs), Medicare cost reports, and demonstration Uniform Billing (UB-92) forms. MPR is collecting information from three primary sources. First, we are conducting telephone interviews with a sample of patients three and eight months after their admission to home health to obtain data on their health and functional status, satisfaction with home health care, and use of non-Medicare services. Second, we have asked demonstration agency nurses and aides to provide data from a single day on the length of their Medicare home health visits. Finally, we are asking agencies to complete an annual self-administered survey, which will provide information on agency characteristics and procedures, including structural measures for the analysis of care quality.

We will use regression analysis or a related multivariate statistical technique to control for exogenous differences that may exist between treatment and control groups despite agencies' random

assignment to treatment or control status. These explanatory variables will be derived from the same source as the outcome measures, as well as from the Section 223 limit files, the Provider of Services File, and the Area Resource File (ARF). For cross-sectional data sets, we will use ordinary least squares regression to estimate models for continuous dependent variables and logit analysis for binary dependent variables. For analyses with agencies as the unit of observation and data for multiple years, we will use fixed- and random-effects regression models. Observations are weighted in the main analyses so that each agency has equal representation. Standard errors are calculated with special software that takes into account the effects of clustering and weighting.

D. HYPOTHESIZED EFFECTS ON QUALITY OF CARE

Prospective payment for health care creates incentives for providers to maximize profit by cutting costs. As with prospective payment to health plans by capitation and to hospitals by diagnosis-related groups, policymakers are concerned that per-episode prospective payment for Medicare home health care may adversely affect quality of care. In response to demonstration incentives, agencies might reduce the number of visits per episode or visit length, hire staff with less experience or training at lower pay, or cut back on staff training, supervision, and clinical documentation.

On the other hand, reactions to the incentives of prospective payment may increase the quality of care. In their efforts to maximize profits, home health agencies might try to deliver more coordinated and effective care so that patients recuperate faster and with fewer visits. Agencies might, for example, implement standardized treatment protocols, recruit wound care specialists, or improve the process for determining when a patient is ready for discharge from home health care and any support they may need then from other sources. By intensively instructing patients and family members in self-care and by involving community-based services providers early on, agencies might

at the same time reduce patients' reliance on home health care and improve patient outcomes while trimming episode length and costs.

Several factors also militate against the likelihood that quality of care will be sacrificed for financial gain. First, in the highly competitive home health market, an agency that develops a reputation for poor quality among referring providers and patients could lose market share, revenue, and thus economies of scale. Second, patient care staff in home health agencies (nurses, therapists, and aides) tend to have a strong personal and professional commitment to providing high-quality care. Thus, even if agency management planned to cut services drastically during the demonstration, it would be difficult to convince patient care staff to do so. Finally, Medicare certification surveys and the independent demonstration quality assurance process provide ongoing monitoring of quality.

Previous studies on the possible effects of reductions in home health visits on quality of care are limited and have reached different conclusions. A comparison of home health use in the fee-for-service and managed care sectors found that HMOs cut back markedly on home health visits per episode and that this reduction may have contributed to poorer functioning among HMO enrollees (Shaughnessy et al. 1995). On the other hand, studies of regional variation in home health use suggest that modest to moderate reductions in home health visits may not lead to poorer patient outcomes (or increased use of other services) (Welch et al. 1996; ProPAC June 1996; and Schore 1995). The wide variation across regions in the number of visits provided suggests that considerable discretion may be involved in planning the amount of home health care for a given patient. Discussions with treatment agency staff at the start of the demonstration supported the notion that some visits could be eliminated without adverse effects on patient care.

Although patient health status and mortality are appealing measures of quality of care, the links between care provided and such patient outcomes are not necessarily straightforward. The

attribution of patient outcomes to quality of care is even more difficult for home health care than it is for hospital or nursing home care. External factors and patient behaviors, such as the patient's ability to afford, purchase, and prepare food; the patient's compliance with medications, diet, or therapy; and the condition of the patient's home, all of which are controlled in an institutional environment, may significantly affect outcomes in the home care setting. The availability and quality of help from family and friends may also have important effects on outcomes. The advantage of the randomized trial design of the demonstration is that such unmeasurable and uncontrollable factors should be in large part balanced between the treatment and control groups, allowing us to attribute treatment-control differences to the effects of the demonstration.

We also use some health service utilization measures as outcomes, but we must be cautious as well in interpreting them as indicators of quality. Hospital admission during the episode, for example, could indicate either poor-quality care by the agency or appropriate referral to a more intense level of care. Emergency visits may result either from inadequate home care or from earlier identification of problems needing urgent attention. Elimination of necessary home health care by agencies may cause deterioration in the conditions of some patients and require their admission to another home health agency. Elimination of unnecessary or discretionary visits, however, may still lead to patient dissatisfaction. Dissatisfied patients might then seek home care from another agency which, in a competitive home health market, may be eager to provide it. Reductions in home health care services might also lead to the substitution of services in other settings (such as SNFs or hospices) without harming quality of care or patient outcomes. Although agencies may benefit financially by discharging costly patients after persuading them (and their families and physicians) that a SNF or hospice is more appropriate, such settings may, in fact, be better for the patient. Thus, in the absence of other corroborating evidence, a finding of different likelihood between patients of

treatment and control agencies of admission to a hospital, a SNF, or another home health agency does not necessarily indicate differences in the quality of care.

E. GUIDE TO THE REST OF THIS REPORT

The second chapter of this report describes the data sources, samples, and specific analytic approaches used in this interim analysis of impacts on quality of home health care for the early period of the demonstration. Chapter III describes findings and summarizes our conclusions.

II. DATA AND METHODOLOGY

Using demonstration QA and claims data, we estimate impacts of the demonstration on patient outcome and health service utilization indicators of the quality of home health care. Depending on the data set analyzed, we report impact estimates either as unadjusted treatment/control means or regression-adjusted means that control for preexisting agency and patient differences.

A. DATA SOURCES AND ANALYSIS SAMPLES

To assess care quality, we analyzed two separate samples of home health care episodes: the QA Sample and the Claims Sample. Data for control variables were drawn from several sources. We describe each sample, data sources, the data collection period, and data items. Table II.1 lists quality indicators, by data source.

1. QA Sample and Data

As the QA contractor, CHPR designed and implemented a patient-outcome-based, quality-monitoring and continuous-improvement system for the demonstration. All agencies participating in the demonstration (in both the treatment and control groups) are required to collect and submit QA information to CHPR. The QA data collection instruments are scaled-down versions of CHPR's full Outcome Assessment System Information Set (OASIS) (Shaughnessy et al. 1995). CHPR calculates agency-level risk-adjusted profiles of patient outcomes from the QA data, and these profiles are then regularly "fed back" to the agencies to help them improve quality of care. (Shaughnessy et al. February 1995).

TABLE II.1

OUTCOME MEASURES FOR QUALITY OF CARE AND DATA SOURCES

Measures	Source
17 Measures of Improvement in Health Status over the Course of the QA Episode[a,b]	Demonstration QA data[c]
17 Measures of Stabilization in Health Status over the Course of the QA Episode[d,b]	Demonstration QA data[c]
Reported Emergency Visit to Hospital Emergency Room	Demonstration QA data[c]
Reported Emergency Visit to Outpatient Clinic or "Urgent Care Center"	Demonstration QA data[c]
Reported Emergency Visit to Physician's Office	Demonstration QA data[c]
Admission to Hospital During "At-Risk" Period[e]	National Claims History File
Admission to SNF During "At-Risk" Period[e]	National Claims History File
Admission to Another Home Health Agency During "At-Risk" Period[e]	National Claims History File
Death During "At-Risk" Period[e]	HCFA's Enrollment database[f]

[a]For example, whether Improved in Pain or Improved in Ambulation.

[b]QA episode refers to the time span between start of care and the follow-up point: discharge or 120 days after admission, whichever occurred first.

[c]Collected by Center for Health Policy Research, University of Colorado, the QA contractor for the demonstration.

[d]Whether Stabilized in the same measures as the Improvement outcomes.

[e]The "at-risk" period refers to the 120-day period following admission to home health for both treatment and control agencies.

[f]Current HCFA repository of enrollment and entitlement data for persons who are or who have ever been Medicare beneficiaries.

A QA start/resumption-of-care instrument is completed by demonstration agency nurses for each patient at start of care or on resumption of care following an inpatient stay of 48 hours or more. A QA follow-up/discharge instrument is completed at discharge or 120 days after admission, whichever comes first. While most demonstration agencies implemented QA data collection in May 1996, a few did not do so until later. Only 87 of the 91 demonstration agencies had collected sufficient QA data for this interim analysis.

a. QA Episodes and QA Analysis Sample

Each observation in the QA data consists of a home health care episode that we call a "QA episode" to differentiate it from a Claims data episode, which, as we describe later, is defined quite differently. To create an observation for a QA episode, we relied on the QA start instrument completed at admission and on its corresponding QA follow-up/discharge instrument completed at discharge or 120 days after admission. Although a QA follow-up form was completed for the last home health visit preceding an inpatient stay, we did not use this form for the patients who returned to the home health agency following the inpatient stay, because they had not been discharged from home health at that time. We also did not use QA start instruments completed at resumption of care, that is, those completed after *return* to home health care following an intervening inpatient stay of 48 hours or more; outcomes of such episodes are difficult to interpret. The start dates of QA episodes defined this way range from early May 1996 to late April 1997.

As the demonstration intervention occurs at the agency level and the agency is the behavioral unit of interest, our main analyses are conducted with observations weighted to represent each agency equally. However, very small agencies could exert undue influence in an analysis with agencies represented equally, and in fact, the number of episodes per agency ranged from 1 to 2,352. We dropped five agencies with fewer than 20 episodes each (54 episodes), leaving 82 agencies (42

treatment and 40 control). We also dropped 91 QA episodes that lasted one day or less. Thus, from the original 28,799 QA episodes, 28,654 remained for analysis.

b. Outcome Variables for QA Data

Health Status Measures. We examined 17 health status measures in the QA data, including Basic and Instrumental ADLs (for example, Bathing, Grooming, Dressing, Eating, Transferring, Management of Oral Medications, Light Meal Preparation), and Physiologic and Psychosocial measures (for example, Pain Interfering with Activity, Dyspnea, Confusion). Measures are scored on an ordinal severity scale. For example, possible responses to the item "how often does pain interfere with the patient's activity/movement" range in severity from "none of the time" to "all of the time." Binary change variables of Improvement or Stabilization in each item are calculated from the admission and follow-up data. Improvement in a measure has the value one if a patient improves in the scale for that item on follow-up, zero otherwise. Patients cannot improve in a measure if they already start the home health episode at the best level of that measure, so such observations are excluded from this measure. Stabilization in a measure has the value one if a patient does not worsen on the scale for that measure on follow-up (that is, remains the same or improves). We exclude patients who start the episode at the worst level, since they cannot worsen on the measure. Thus, the number of observations used differed for the Improvement and Stabilization outcomes and differed across functioning measures.

Reported Emergency Health Services Use. Based on reports by patient, family, or other providers, agency nurses recorded on the QA follow-up/discharge instrument any emergency visits to hospital emergency rooms, physician offices, outpatient clinics, or freestanding urgent care centers made since start of care.

c. Limitations of QA Data for Demonstration Evaluation

One feature of the QA data that makes comparisons of outcomes between the treatment and control group difficult to interpret is the endogenous timing of the measurement of QA outcomes. The QA data collection system was designed for home health care continuous quality improvement, not for evaluation of the per-episode payment demonstration. A QA follow-up/discharge instrument is completed at discharge or at 120 days after admission, whichever occurs first (Shaughnessy et al. 1994). Discharge from home health care is the most frequent trigger for completion of the QA follow-up instrument, but in treatment agencies demonstration incentives directly encourage earlier discharge and shorter episode. Thus data for 13.4 percent of control group cases were collected at 120 days after admission, whereas only 9.9 percent of treatment group cases were collected at this point. The weighted mean QA episode length, defined as the date of the QA follow-up/discharge instrument minus the date of the QA start instrument, was 45 days for treatment agencies versus 58 days for control agencies ($p<0.0001$, two-tailed test).[1] Compared to patients of control agencies, patients in treatment agencies thus have their outcomes measured sooner in their home health episode. Depending on the shape of the "natural progression of patient condition" for each health status measure, patients of the treatment agencies may thus appear to be either worse or better off simply because their outcomes are measured at an earlier time point, not because of any differences in the content or quality of care (Phillips et al. 1995).

[1]These mean values are calculated with episodes weighted to represent agencies equally. The significance test for equality of the weighted mean values of Time Under Observation between treatment and control groups does not account for design effects due to clustering. There is no need to account for design effects in the comparison here because we are interested only in differences within this sample, not the universe of all agencies. This p-value does include, however, the effects of weighting on the variance.

The quality of the QA follow-up data may also be different for the treatment and control groups because it may vary with the frequency of home health visits. Agency staff depend on patient or family reports to learn of such occurrences as behavioral problems or visits for emergency care. Ascertainment of these outcomes likely varies depending on how often staff visit the patient and family, yet the demonstration incentives for treatment agencies are to reduce numbers of visits. In short, it is difficult to disentangle independent demonstration impacts on QA outcomes from those that stem from demonstration impacts on the endogenous variables of episode length and visit frequency.

By measuring patient outcomes at the same fixed time points after home health admission in both treatment and control agencies, the demonstration patient survey fielded by MPR avoids the problems of the QA data. The survey collects functional status and satisfaction data from patients three and eight months after home health admission. Patient survey data are not available for this interim report but will be the focus of the final report.

2. Medicare Claims Sample and Data

A different sample of home health episodes was constructed from Medicare claims data for patients admitted to each agency between the agency's demonstration start date and August 31, 1996--the first 8 to 15 months of the demonstration.

a. Claims Episodes in Claims Data

UB-92 bill record files obtained from the demonstration FI, PGBA, were used to identify home health episodes in the Claims data (Claims episodes) as defined by demonstration rules for patients of both treatment and control agencies. Beginning with each agency's enrollment in the demonstration, we scanned the UB-92 files to identify the first admission and subsequent bill records

for each individual. To create a Claims episode, we combined all records for an individual for 120 days following the first admission and any bills for care after 120 days until we observed a gap of at least 45 days in billing dates. This procedure was followed regardless of whether the agency discharged and readmitted a patient during the 165 (120 + 45) days. If we observed another home health admission after 165 days (preceded by a 45-day or longer gap in care), we created a second episode for that individual, and so on for any subsequent episodes beginning through August 31, 1996. The Claims episodes as defined here are the same as the home health episodes analyzed in the interim reports on home health use (Cheh et al. 1997) and use and cost of Medicare services (Schore 1997) but are different from the QA episodes in the QA data defined above in Section A.1.a.

b. Outcome Variables for Claims Data

We merged the episodes identified from UB-92 bill record data to HCFA's Medicare standard analytic files (SAF). The HCFA National Claims History SAF was used to identify Medicare Part A health service use (inpatient hospital stay, admission to SNF, admission to another home health agency) occurring in the 120-day period for each episode identified from the UB-92 files. We also used death dates in HCFA's enrollment database to construct an indicator of whether a patient died within the 120-day period. Matches were identified for 57,261 episodes; only 493 episodes (less than one percent) did not match. Medicare (Part A) service-use data from the SAF were used to measure service receipt during the 120 days after home health admission. We constructed indicators of whether a patient was admitted to a hospital, a SNF, or another home health agency within the 120-day period for a diagnosis involving the same body system as the condition for which the patient was originally admitted to the demonstration agency (which we call "same-body-system admissions"). The ICD-9 coding manual groups ICD-9 diagnosis codes into 13 body systems (for example, Diseases of the Circulatory System). A hospital admission, SNF admission, or home

health readmission within the 120-day period was counted as an outcome only if the principal or first additional ICD-9 diagnosis for that admission fell into the same body system as the principal or first additional ICD-9 diagnosis for the original home health admission.

Medicare claims were extracted from the SAF in May 1997. Since claims are generally included in the SAF within four months after the service was rendered, we should have nearly complete data on Part A services received through December 1996.

c. Control Variables for Claims Data

We controlled in the claims sample analysis for several patient-, agency-, and area-level variables that we considered important potential confounders of the patient outcome measures. Table II.2 summarizes the control variables used in the claims data analysis and the sources of the variables.

Patient Characteristics. Control variables for patient characteristics were used in the analysis to account for possible differences in patient mix between treatment and control agencies. Home health patients who are more severely ill or more functionally impaired or who carry certain diagnoses are more likely to be admitted to a hospital, a SNF, or another home health care agency and are more likely to die in the 120-day follow-up period than beneficiaries without these health problems.

We obtained data from three sources on patient characteristics at the start of the home health episode: (1) the remarks field of the UB-92, used by the demonstration's case-mix adjuster; (2) Medicare enrollment databases; and (3) the Medicare SAF. In the remarks field for the first UB-92 bill following a demonstration admission, both treatment and control agencies were required to submit the information on patient characteristics needed for the 18-category Home Health Utilization Group (HHUG) case-mix adjuster. The characteristics include measures of impairment in

TABLE II.2

STANDARD CONTROL VARIABLES FOR MULTIVARIATE ANALYSIS OF CLAIMS DATA, BY SOURCE

Episode Level		Agency Level		Area Level
Data Source/Patient Characteristics at Episode Start	Medicare Service Use in Year Preceding Episode (Medicare Standard Analytic Files)	Base-Quarter Patient Service Use (Demonstration Case-Mix File)	Agency Characteristics	County Characteristics (Area Resource File)
UB-92 Remarks	Length of inpatient stay during 2 weeks before home health	Agency Predemonstration Practice Pattern	Base Year Cost Reports	Physicians per 10,000 residents (1994)
Has cancer	Whether in SNF during 2 weeks before episode start		Was hospital-based	Nursing home beds per 100 elderly residents (1991)
Has diabetes	Total Part A Medicare reimbursement in the 6 months prior to episode start		Profit status	Hospital occupancy rate (1993)
Had stroke	Total number of Part A home health visits in the 6 months prior to episode start		Agency size	Mean Medicare reimbursement per beneficiary (1991)
Has decubiti stage 3 or 4	Whether hospitalized in the 6 months prior to episode start		State	
Needs complex wound care			Demonstration Implementation Contractor	
Has limitations in bathing, eating, dressing, toileting, transferring			Urban/rural indicator	
Admitted to home health from hospital			Chain membership	
Medicare Enrollment Database				
Age				
Gender				
Race				
Original reason for entitlement				
Had Medicare for less than 6 months				

Activities of Daily Living (ADL) and whether the patient has certain medical conditions (cancer, diabetes, decubitus ulcers) and care needs (complex wound care). Medicare enrollment files provide us with basic patient demographic information, including the patient's age (at the start of home health episode), gender, race, and disability status (from the original reason for Medicare qualification). From the SAF, we constructed measures of Medicare service use to capture the patients' severity of illness, including measures of recent acute illness (whether admitted from hospital, length of prior hospital stay) and longer-term home health use (six months prior to admission). For the latter, we used the mean value for beneficiaries between 65.5 and 66 years old as a proxy measure for beneficiaries age 65 to 65.5 at home health admission, since beneficiaries in this age range would not have been eligible for Medicare service for a full six months prior to their home health admission.

Agency and Area Baseline Characteristics. Different types of agencies may have different goals, practice styles, and cost and management structures that could affect the home health care they render. In addition, patient case-mix may vary by type of agency. For example, proprietary and nonprofit agencies might have different preexisting practice patterns in the number of visits rendered per episode, and hospital-based agencies might serve a higher percentage of patients with acute (rather than chronic) conditions than freestanding agencies.

Agency characteristics are also used to define subgroups, because agencies with different goals, cost and management structures, and practice patterns could respond differently to the incentives of the demonstration. For example, proprietary agencies may have a stronger interest in revenue surpluses (profits) than nonprofit agencies, reduce visits by a greater margin, and affect quality of care to a greater degree.

Data on agency characteristics were obtained from base-year Medicare cost reports and from the demonstration implementation contractor, Abt Associates. Base-year Medicare cost reports provided information on the agencies' base-year characteristics, including for-profit status, affiliation, and size (as measured by total number of visits rendered). Information Abt Associates gathered during demonstration recruitment included agency chain membership and rural location.

We also constructed an agency-level variable reflecting each agency's practice pattern during its predemonstration base quarter. The variable is a case-mix-adjusted ratio of the average number of visits received by each agency's patients in the 120 days after admission during its predemonstration base quarter to the average number provided by other demonstration agencies. The variable was calculated from case-mix data Abt Associates collected for each agency's predemonstration base quarter. For each agency, we first computed ratios of the average number of visits (of each type) the agency delivered to patients in each case-mix cell to the average for all agencies for that case-mix cell.[2] Within each cell, we constructed a weighted average of the ratios for the six visit types. The final index was the weighted average of the ratios in the 18 case-mix cells. The weight for a given cell ratio was the agency's proportion of episodes in that cell in the base quarter. A practice-pattern index value greater than one indicates that, controlling for differences in case mix, an agency provided more visits during the 120-day period than did other demonstration agencies during the quarter preceding the demonstration. The weighted mean predemonstration practice pattern index for treatment agencies was about 15 percent lower than for control group agencies (0.93 versus 1.10, $p<0.01$, two-tailed test).[3] Inclusion of this explanatory

[2]The 18 cells of the Demonstration case-mix adjuster, HHUGs.

[3]For a discussion of how weighting and clustering affect the p-value for this comparison, see Footnote 1 above.

variable in the models allows us to control for the preexisting differences between the treatment and control agencies in propensity to provide services, which may be related to predemonstration differences in average patient outcomes. Thus, similar differences in outcomes occurring during the demonstration will not be attributed to program effects.

We also controlled explicitly for area-level characteristics that might influence the outcomes under study. For example, likelihood of admission to a SNF may be lower (or to another home health care agency, higher) in areas where the supply of nursing home beds is limited relative to demand. We obtained area characteristics from the ARF, including physicians per 10,000 residents, nursing home beds per 100 elderly residents, and hospital occupancy rates.

d. Summary Statistics for Control Variables in Claims Data

Table II.3 displays the treatment and control group means for the explanatory variables in our regression models. (Following the methodology described in Section B.4 of this chapter, these means have been constructed using sample weights that give each agency equal representation.)[4]

With the large sample of episodes that we have available, we have statistical power to detect very small differences between the treatment and control groups at baseline.[5] We therefore expect

[4]The significance levels for the tests of equality between treatment and control group means in Table II.3 do not account for design effects due to the clustering. For our purposes here (but not in the main analysis described in Section B.3), there is no need to account for these effects, because we are interested only in differences within this sample, not the population of all agencies. We do account for the design effects associated with our use of sample weights, however. See Section B.3 of this chapter for a complete discussion of the use of weighting and clustering in our analysis.

[5]For a binary indicator with a mean of 50 percent, for example, the minimum detectable difference is 1.3 percentage points, using a two-tailed test with a 95 percent confidence interval, at 80 percent power.

TABLE II.3

WEIGHTED MEANS FOR STANDARD BENEFICIARY-SPECIFIC CONTROL VARIABLES, BY TREATMENT STATUS
(Percentage, Unless Otherwise Noted)

	Treatment Group	Control Group
Age		
Less than 65	7.9	9.0**
75 to 84	39.7	40.0
85 or older	23.5	21.7**
Female	63.3	64.4
White	80.9	81.1
Original Reason for Medicare: Old Age	83.3	81.4***
Medical Conditions		
Cancer	12.7	12.8
Diabetes	21.6	21.7
Cerebrovascular accident (stroke)	15.3	14.7
Decubiti stage 3 or 4	4.6	3.7***
Need for Complicated Wound Care[a]	7.0	7.0
Functional Limitations[b]		
Bathing	71.8	72.7
Eating	28.4	30.2**
Dressing	60.4	64.6***
Toileting	38.1	40.6***
Transferring	49.7	52.2***
Preadmission location: hospital	35.2	37.9***
Had Medicare for less than 6 months	1.4	1.4
Length of hospital stays ending during 2 weeks before home health admission (days)[c]	3.6	4.2***
Any SNF stay during 2 weeks before home health	17.4	15.4***
Total Medicare Part A reimbursement during 6 months before home health (in 1,000s of dollars)[d]	11.4	11.4
Number of Episodes	26,281	25,032

TABLE II.3 *(continued)*

SOURCE: Medicare Enrollment Files and demonstration-specific fields from UB-92 forms.

[a]Patient has wound that requires soaking, irrigation, or debridement.

[b]Patient requires some human assistance with or does not participate in activity.

[c]If patient was not hospitalized withing the two weeks before home health, days are set to zero.

[d]Includes reimbursement for inpatient, SNF, hospice, and nondemonstration home health paid under Medicare Parts A and B.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.

that many differences in the explanatory variables between the treatment and control agencies will be statistically significant even when the magnitude of the difference is not large.

Demographic Measures. Some of the demographic characteristics of patients admitted to treatment and control agencies differed significantly, but the magnitudes of the differences are not large. Relative to control agencies, treatment agencies served a slightly higher proportion of patients age 85 and older and about two percent more patients originally entitled to Medicare because of age.

Medical Conditions and Care. There was only one statistically significant difference between treatment and control agencies in the means for measures of medical conditions and care needs drawn from the case-mix data in the remarks field of the UB-92: treatment agencies were significantly more likely to serve patients who had a decubitus ulcer. While the percentage is low for both groups and differs by less than one percentage point, the treatment group mean is about one-fourth larger than the control mean.

Limitations in ADLs. Relative to control agencies, treatment agencies served patients who were less likely to be impaired in ADLs, but the magnitudes of the differences are quite small. For three of the six ADL tasks for which we have measures (dressing, toileting, and transferring), treatment agencies had significantly lower sample means than control agencies. All these treatment-control differences are less than 10 percent of the control group mean.

Patient Prior Service Use Measures. Patients served by treatment and control agencies differed somewhat in their use of medical services prior to entering the index home health episode. Treatment agencies served patients who were more likely to have been recently discharged from a SNF and less likely to have been admitted after release from a hospital. These differences offset one another: each is roughly two-and-a-half percentage points. In addition, prior hospital stays were shorter for patients in the treatment group than in the control group. For the treatment group, the

mean length for a prior hospital stay was about half a day (14 percent) shorter than for the control group.

Agency Characteristics. Treatment and control agencies differ significantly on nearly all the agency and area characteristics examined (Table II.4), and a few of these differences are large. While Table II.4 compares the percentages of treatment and control episodes with the specified agency and area characteristics, the use of weights to represent agencies equally allows us to compare treatment and control agencies. Treatment agencies are significantly less likely than control agencies to be proprietary, this difference is not large. On the other hand, compared to control agencies, treatment agencies are substantially (1) less often hospital-based, (2) more often affiliated with a chain, and (3) smaller. As previously noted, we also observe a substantial and statistically significant preexisting treatment-control difference in practice patterns. On average, the predemonstration practice pattern index for the treatment agencies is about 15 percent lower for treatment group than for control group agencies.

Area Characteristics Measures. Significant differences also exist between treatment and control agencies in area characteristics (Table II.4, bottom). The distribution across states differs for the treatment and control agencies, with the treatment group overrepresented in Massachusetts and underrepresented in Illinois relative to the control group. Given the wide differences across regions in practice patterns, this difference could be important. There are also significant differences in the rate of physicians per 10,000 residents and the hospital occupancy rate; however, both of these differences are very small, representing less than a three percent difference relative to the control group mean.

In summary, while the preexisting treatment-control differences in patient characteristics are minimal, there are several large differences in agency and area (state) characteristics. To the extent

TABLE II.4

WEIGHTED MEANS FOR STANDARD AGENCY- AND AREA-SPECIFIC CONTROL VARIABLES, BY TREATMENT STATUS
(Percentages of Episodes, Unless Otherwise Noted)

	Treatment Group	Control Group
Agency Characteristics		
Predemonstration Ratio of Mean Agency Visits to Mean for All Demonstration Agencies (Case-Mix Adjusted)	0.93	1.10***
Hospital-Based Agency	8.7	14.6***
For-Profit Agency	47.8	51.2***
Chain Member	37.0	26.8***
Agency Provided Fewer than 30,000 Visits in Base Year	34.8	19.5***
Area Characteristics		
Agency Located in Urban Area	84.8	85.4
State		
Florida	8.7	9.8**
Illinois	13.0	22.0***
Massachusetts	17.4	7.3***
Texas	34.8	39.0***
County-Level Means		
Number of nursing home beds per 100 persons over age 65	5.08	5.16*
Number of physicians per 10,000 persons	22.00	21.48***
Hospital occupancy rate	0.62	0.61***
Medicare reimbursement per beneficiary (in thousands of dollars)	3,413	3,406
Number of Episodes	**26,282**	**25,031**

SOURCE: Medicare Cost Reports; Area Resource File.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

that these characteristics affect the patient outcomes we use to measure quality of home health services, we would incorrectly estimate the effects of per-episode payment with simple comparisons of treatment and control group means. Regression models allow us to control for these differences and provide more accurate estimates of the effect of the demonstration payment method.

e. Claims Analysis Sample

From the initial set of 91 agencies, one treatment group agency was dropped because it had no admissions and three agencies were dropped because they withdrew from the demonstration after participating for only a few months, leaving 87 agencies (46 treatments, 41 controls). The three agencies that dropped out of the demonstration early either were purchased by another agency or merged with another agency, and the new ownership did not want to be part of the demonstration. Loss of these agencies accounted for 1,613 Claims episodes.

We dropped 1,114 Claims episodes because we learned from the HCFA enrollment files that these patients were enrolled in a Medicare HMO at the time of their home health admission and thus not eligible to participate in the demonstration. The admitting demonstration agencies, probably unaware that these patients were HMO members when they admitted them, submitted a claim to the demonstration FI in error. We excluded another 1,267 Claims episodes because Medicare was a secondary payer. Agencies were not entitled to a per-episode payment for these episodes. Treatment and control agencies are both reimbursed at cost during the demonstration for care for which Medicare is a secondary payer.

Finally, we excluded 1,735 Claims episodes that were missing the patient characteristic data from the remarks section of the UB-92. These data had been inadvertently erased from the UB-92s because of an error in the software used by PGBA, the demonstration FI. PGBA is working to restore the data but could not complete the task in time for this report.

For the remaining episodes, we constructed from the various data sources an analysis file in which each observation contains data on (1) the patient's use of Medicare Part A services during the home health episode; (2) the patient's vital status at 120 days after admission (dead or alive); (3) the patient's characteristics at admission to home health; (4) the patient's Medicare service use before the home health admission; (5) the predemonstration characteristics of the home health agency providing care, and (6) the characteristics of the area in which the agency is located. Of the 57,216 Claims episodes initially constructed, 51,313 episodes remained in the analysis sample.[6,7]

f. Claims Data for Demonstration Evaluation

The utilization and mortality outcomes of the Claims data lack the richness of the QA data's functional and health status outcomes. However, we can be more confident that any treatment-control differences we find in the Claims data are truly due to demonstration impacts since, unlike the QA outcomes, outcomes in the Claims data are comparably measured for the two groups.

B. METHODS

1. QA Data

Following our original analysis plan, we present only comparisons of unadjusted treatment and control group means in this interim report. Because the QA sample and the Claims sample are different samples from different periods of time, the extensive Claims data control variables are unavailable in the QA data. Linking the QA data with the Claims data for patient-level control

---

[6]Some episodes were excluded for multiple reasons, so the numbers presented do not sum to 5,903, the difference between 57,216 and 51,313.

[7]There is a difference of one case between the analysis file for the report on Medicare Service Use and Reimbursement (Schore 1997) and our analysis file for Quality of Care impacts (that is, 51,314 vs. 51,313). A case initially dropped from both analysis files because of a negative Medicare cost figure was later restored only to the Medicare Service Use analysis file.

variables would result in the loss of large numbers of cases, as the result of the differing time periods and the suspect quality of patient identification numbers. Agency and area variables in the Claims data are also for a time period different from that of the QA data. However, since the QA outcome variables of Improvement and Stabilization are measures of *change* in health status, adjustment for patients' baseline levels of impairment is less important in their interpretation. In the final demonstration impact report on quality of care, we will have control variables available for the QA data.

Since the outcome variables in the QA data are all binary (for example, whether a patient experienced improvement in a health status measure over an episode, whether an emergency outpatient visit occurred during an episode), we use the chi-square statistic to test whether the outcomes are significantly different between treatment and control groups.

2. Claims Data

For the Claims data analyses, we used multiple logistic regression to estimate treatment-control differences in the binary dependent variables while controlling for other important variables. The structure of the logit model is as follows:

$$(1) \quad Probability\ (Y=1) = \frac{1}{1 + e^{(\alpha + X\beta + \delta T)}},$$

where Y is the (binary) outcome variable and the remaining variables and parameters are as follows:

X is a vector of control variables

T is a binary variable for treatment status that equals 1 for episodes rendered by treatment agencies and 0 for episodes rendered by control agencies

α is the intercept term

β is the vector of regression coefficients on the control variables

δ is a parameter that measures the impact of prospective rate setting on the outcome Y

ϵ is a random disturbance term assumed to have a mean of zero (conditional on X and T) that reflects all the unobserved factors affecting Y

Given the nonlinearity of the logit model, the estimated impact of the payment method is not measured directly by the coefficient δ on the variable for treatment status. To estimate the demonstration impact on the probability that $Y = 1$, we use the coefficient estimates from the model to generate two predicted probabilities for each observation: one assuming that the observation belongs to the treatment group ($T = 1$), and one assuming that it belongs to the control group ($T = 0$). The impact estimate is the average difference between these estimated probabilities. Because the statistical significance of δ determines whether the odds that $Y = 1$ are significantly different for the treatment and control groups, we use the p-value for this parameter to test our hypotheses about differences between the two groups.

Throughout the tables of results for the Claims data, we present as a point of reference the unadjusted control group mean of each outcome variable alongside the estimated impact. The unadjusted control group mean provides a reasonable estimate of the mean value for the outcome variable that might be expected to occur in the absence of the demonstration. We use this mean to assess the relative magnitude and importance of the estimated impact.

3. Hypothesis Tests for the Impact Estimates

For both the comparisons of means and the logit analyses, two-tailed statistics are used to test the null hypothesis that there is no difference in outcomes between treatment and control agencies. The p-values are based on estimated standard errors that account for the clustering of episodes within agencies and the use of sample weights. A p-value below 0.10 indicates rejection of the null

hypothesis and provides significant statistical evidence that a demonstration impact exists. At this p-value, however, approximately 10 percent of independent tests will show, simply by chance, a statistically significant treatment-control difference when there is no true program effect (known as Type I error). Therefore, in assessing whether a statistically significant treatment-control difference, especially one with a p-value between .05 and .10, should be interpreted as a true program impact, we consider whether the sign and magnitude of the predicted effect are consistent with those for related outcomes.

Despite our large sample of patient episodes, it is unlikely that we will be able to identify small demonstration impacts with our sample, because design effects greatly reduce the precision of our estimates. For example, ignoring the design effects associated with our weighted sample and the clustering of episodes within agencies, the minimum detectable effect of the demonstration on the binary outcome measure for improvement in pain (which has a mean of about 50 percent) is about 1.5 percentage points, under a two-tailed test at the 10 percent significance level, with 80 percent power. After accounting for design effects, however, the minimum detectable effect is about 7.9 percentage points. Thus, despite our large sample sizes, it is unlikely that we would find significant effects of prospective payment on outcomes unless they were at least moderate in size.

4. Weighting

As noted previously, we weight the episodes in the main analyses to give agencies equal representation in the analysis. We use this approach for two reasons. First, because the demonstration is implemented at the agency level (not the episode level), the agency is the behavioral unit of interest. Second, the use of weighted data ensures that the impact estimates will not be dominated by the experiences of a few large agencies.

For each agency *i*, we construct the "agency equal" weight as follows:

(2) $w_i = \frac{1/n_i}{k/n}$,

where n_i is the number of (episode-level) observations in agency i, k is the number of agencies, and n is the total number of observations for all agencies. In the QA data, the weights range from 0.15 to 17.5, with 75 percent of the agencies having weights between 0.36 and 1.03. In the Claims data, the weights range from 0.14 to 25.4, with 75 percent of the agencies having weights between 0.31 and 1.13.

Weighting observations so that agencies are represented equally may accord small agencies undue influence in the analysis. Although we dropped five agencies from the analysis of the QA data because they had fewer than 20 episodes, there were still 14 agencies with less than 60 episodes each. To investigate the sensitivity of our results to the weighting approach, we also present the impacts of the demonstration when each agency is represented in the analysis proportional to its size, as measured by its share of demonstration episodes. Similar results from the two approaches would suggest strongly that small anomalous agencies did not exert undue influence and that the results may be broadly interpreted for policy purposes. Conversely, while dissimilar results under the two sample weighting schemes do not necessarily indicate that the main results are incorrect, they indicate that further analysis is needed to draw appropriate inferences about the likely effects of a national program.

Since all episodes for each agency are included in the analysis, representing each agency proportional to its size would ordinarily mean conducting the analysis without sample weights. For this interim report, however, our samples include home health admissions occurring over different lengths of time for different agencies.[8] Thus, in order to reflect the agencies' relative size

[8]We currently have data over varying time intervals because agencies entered the demonstration (continued...)

accurately, we must scale each observation for an agency by the (relative) time that it had been in the demonstration at the time the file was cut off (August 1996 for the claims sample and April 1997 for the QA sample).

For each agency *i*, the "agency share" weight that reflects its time in the demonstration is

$$(3) \quad w_i^s = \frac{\bar{t}}{t_i},$$

where t_i is the length of time that agency *i* has been in the demonstration as of the cutoff date for the sample being analyzed, and $\bar{t}$ is the average of the t_i's across agencies. The values of t_i and $\bar{t}$ differ between the QA sample and the Claims sample because of the different sample time frames and varying agency start-up dates for QA data collection.

5. Design Effects

Correct statistical testing of hypotheses depends on proper estimation of standard errors. The observations in our data are not independent, because they are clustered within the demonstration agencies. We also weight observations to represent agencies equally (or alternatively, proportional to their size). Standard statistical software packages that treat observations as a simple random sample of independent observations and do not account for clustering and weighting would seriously underestimate standard errors.

We use the SUDAAN software to obtain the correct standard errors for our impact estimates. The SUDAAN calculations assume that the agencies were selected at random first, then that patients were selected at random from each agency. Neither assumption is entirely true, but the calculations

[8](...continued)
at the start of their fiscal year, which differs by agency, and there was some variation in start-up of QA data collection. For the final report, we will have three years of data on all agencies.

do give us the correct variance that results from selecting episodes for the study in two stages rather than a one-stage random sample of episodes from a universe of all possible episodes.[9] The SUDAAN calculations also account for the greater variance introduced by using sample weights in the regression models.

[9]Our sample actually includes the entire population of patient episodes taking place in demonstration agencies over the early demonstration period. However, because we wish to make inferences about the outcomes for patients admitted in other times and to other agencies, we treat episodes in the data as though they were drawn from the pool of all (future) episodes in all agencies.

III. IMPACTS ON THE QUALITY OF CARE

The interim report by Cheh et al. (1997) demonstrates that per-episode payment has markedly reduced the number of visits home health agencies provided during the first 120 days after admission. We find no evidence that the reduction in visits has adversely affected the quality of home health care. For the most part, differences between treatment and control agencies in patient outcome and health services utilization indicators of quality of care were small and statistically insignificant.

A. QUALITY ASSURANCE DATA

1. Improvement and Stabilization Outcomes

Improvement Outcomes. Table III.1 shows mean rates of Improvement outcomes for treatment and control agencies and unadjusted differences in the means (observations weighted to represent agencies equally). The only two outcome comparisons that reach statistical significance both favor the treatment group. The mean Improvement rate on Confusion was 42 percent in treatment agencies versus 34 percent in control agencies ($p=0.04$), and in Urinary Incontinence or Catheter Present, the mean Improvement rate was 46 percent in treatment agencies versus 38 percent in control agencies ($p=0.09$). Out of 17 comparisons, however, we would expect about one to be statistically significant at the .05 level and two at the .10 level, strictly by chance.

Not surprisingly, short term Improvement rates were higher (80 to 90 percent) in both groups for well-defined problems that are more treatable or likely to improve, such as Pressure Ulcer Count, Most Problematic Pressure Ulcer, and Surgical Wound Status. Rates were lower (30 to 40 percent) for complex problems that are less treatable or likely to improve, such as Ambulation/Mobility, Management of Oral Medications, and Housekeeping.

TABLE III.1

WEIGHTED MEANS OF IMPROVEMENT OUTCOMES
FROM QUALITY ASSURANCE DATA

Improvement in:	Treatment Group (Percentage)[a]	Control Group (Percentage)[a]	Treatment-Control Difference (Percentage)	P-value[b]
Pain	53.6	50.7	2.9	0.36
Pressure Ulcer Count	90.6	90.7	-0.1	0.98
Most Problematic Pressure Ulcer	89.0	88.6	0.4	0.94
Surgical Wound Status	84.0	83.4	0.6	0.81
Dyspnea	47.5	43.3	4.2	0.14
Urinary Tract Infection	79.0	74.0	5.0	0.36
Urinary Incontinence or Catheter Present	45.6	38.3	7.3*	0.09
Confusion	42.0	34.2	7.8**	0.04
Behavior Problem Frequency	65.7	58.8	6.9	0.18
Grooming	50.1	49.7	0.4	0.87
Bathing	50.0	49.1	0.9	0.77
Toileting	54.3	51.7	2.6	0.39
Transferring	47.5	44.0	3.5	0.25
Ambulation/Mobility	32.3	31.8	0.5	0.81
Light Meal Preparation	43.8	45.1	-1.3	0.65
Housekeeping	40.5	40.6	-0.1	0.97
Management of Oral Medications	35.8	35.3	0.5	0.85

SOURCE: Demonstration Quality Assurance Data, CHPR.

[a]Episodes are weighted to give each agency equal representation. The "agency equal" weight for all episodes delivered by agency i is $w_i = \frac{1/n_i}{k/n}$, where n_i is the number of (episode-level) observations in agency i, k is the number of agencies, and n is the total number of observations for all agencies.

[b]The p-values are from chi-square tests based on standard errors inflated to account for clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.

Stabilization Outcomes. Stabilization rates are shown in Table III.2. Control group rates of Stabilization ranged from 76 percent for Housekeeping to 99 percent for Pressure Ulcer Count and Most Problematic Pressure Ulcer. Treatment-control differences were minimal, ranging from -2.1 percent to 2.7 percent (Housekeeping). Only for Stabilization in Behavior Problem Frequency (treatment-control difference of 2.1 percent) and Transferring (treatment-control difference of -2.0 percent) did the comparisons reach statistical significance (p-values for both ≤ 0.05). However, the effects on these two outcomes were in opposite directions

2. Reported Use of Emergency Health Services

Treatment group patients were significantly less likely to report Use of Emergency Health Services in the QA data (Table III.3). Roughly 16 percent of control group patients reported any emergency visit during the period under observation, most of which were visits to hospital emergency rooms (13 percent ER visit rate in control agency group). For both of these measures, the treatment-control difference was statistically significant (-3.5 percent, p=0.01 for the combined measure; -2.8 percent, p=0.02 for hospital ER visits). Emergency visits to outpatient clinics and physician offices were few, and the demonstration had minimal and insignificant effects on these.

3. Ambiguities in the Interpretation of QA Results

The systematic measurement differences between treatment and control groups means the QA results can be interpreted in different ways, depending on the assumptions one makes. As discussed in Chapter II, treatment agencies discharge their patients from home health care more quickly than control agencies, and thus measure their patients' outcomes earlier in the course of home care. Treatment agencies are also likely to measure differently outcomes that depend on patient and family reports to agency staff, because they provide fewer visits than control agencies. In the case of

TABLE II1.2

WEIGHTED MEANS OF STABILIZATION OUTCOMES
FROM QUALITY ASSURANCE DATA

Stabilization	Treatment Group (Percentage)[a]	Control Group (Percentage)[a]	Treatment-Control Difference (Percentage)	P-value[b]
Pain	82.6	83.3	0.7	0.68
Pressure Ulcer Count	99.0	99.1	0.1	0.69
Most Problematic Pressure Ulcer	99.0	99.1	0.1	0.60
Surgical Wound Status	98.2	97.3	-0.9	0.15
Dyspnea	78.5	79.5	1.0	0.47
Urinary Tract Infection	98.1	98.2	0.1	0.70
Urinary Incontinence or Catheter Present	93.4	94.3	0.9	0.36
Confusion	84.0	85.6	1.6	0.23
Behavior Problem Frequency	90.5	92.6	2.1**	0.05
Grooming	86.5	86.6	0.1	0.97
Bathing	82.9	83.2	0.3	0.85
Toileting	91.2	89.1	-2.1	0.25
Transferring	88.8	86.8	-2.0**	0.04
Ambulation	90.2	89.3	-0.9	0.37
Light Meal Preparation	89.2	88.6	-0.6	0.45
Housekeeping	73.3	76.0	2.7	0.72
Management of Oral Medications	87.5	86.4	-1.1	0.33

SOURCE: Demonstration Quality Assurance Data, CHPR.

[a]Episodes are weighted to give each agency equal representation. The "agency equal" weight for all episodes delivered by agency i is $w_i = \frac{1/n_i}{k/n}$ where n_i is the number of (episode-level) observations in agency i, k is the number of agencies, and n is the total number of observations for all agencies.

[b]The p-values are from chi-square tests based on standard errors inflated to account for clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

TABLE III.3

REPORTED EMERGENCY HEALTH SERVICES USE

Reported Emergency Visits to:	Treatment Group (Percentage)[a]	Control Group (Percentage)[a]	Treatment-Control Difference (Percentage)	P-value[b]
Hospital Emergency Room	9.8	12.6	-2.8**	0.02
Hospital Outpatient Clinic	1.0	1.5	-0.5	0.13
Physician Office	2.4	3.0	-0.6	0.26
Any of Above	12.6	16.1	-3.5**	0.01

SOURCE: Demonstration Quality Assurance Data, CHPR.

[a]Episodes are weighted to give each agency equal representation. The "agency equal" weight for all episodes delivered by agency i is $w_i = \frac{1/n_i}{k/n}$ where n_i is the number of (episode-level) observations in agency i, k is the number of agencies, and n is the total number of observations for all agencies.

[b]The p-values are from chi-square tests based on standard errors inflated to account for clustering and weighting.

**Significantly different from zero at the .05 level, two-tailed test.

reported emergency health services use, agency staff visit frequency may influence not only the ascertainment of outcomes but even their occurrence, if agency staff tend to discover problems on home visits and send patients to emergency rooms. Whether or not differences in outcomes are due to the effects of prospective payment depends on the underlying shapes of the "natural progression of patient condition" for the various outcome measures, and on whether home health visit frequency influences event frequency and patient and family reporting of events.

As an illustration, a significant difference in outcome favoring treatment agencies (for example, Improvement in Urinary Incontinence or Catheter Present, Improvement in Confusion, Stabilization in Behavior Problem Frequency, Reported Hospital ER Visit, Any Reported Emergency Visit) is consistent with any of the following interpretations: (1) treatment agencies perform better than control agencies for the outcome measured, (2) treatment agencies perform no better or worse than control agencies but merely appear better because their patient outcomes are measured earlier and the shape of the natural progression curve is downward sloping, (3) treatment agencies perform no better or worse than control agencies but merely appear better because they visit patients less often.[1] There is not enough information available in the data to confirm any particular explanation, and more than one explanation may be operating. We can likewise generate different interpretations for outcomes in which treatment agencies appear significantly worse (for example, Stabilization in Transferring) and outcomes in which there are no significant treatment-control differences (most Improvement and Stabilization outcomes).

[1]The last explanation, for example, is consistent with the findings that reported use of any emergency visit was significantly lower for the treatment group in the QA data, yet there was no significant difference in any emergency room visit or any physician office visit using Medicare claims data (Schore 1997).

B. CLAIMS DATA

In the Claims sample, we estimated the effect of prospective payment on the following events occurring within the 120-day period after home health admission: hospital admission for a diagnosis involving the same body system, admission to a SNF for a diagnosis involving the same body system, admission to another home health agency for a diagnosis involving the same body system, and death. The estimates indicate that there were no program impacts on any of these outcomes (Table III.4).

We observed no demonstration effect on hospital admission for a same-body-system diagnosis during the 120 days after admission to home health. Roughly 19 percent of control group patients had a hospitalization for a same-body-system diagnosis, and the demonstration had negligible and insignificant effect (0.1 percent). Likewise, the demonstration had little effect on rates of admission to a SNF for a diagnosis involving the same body system or on admission to another home care agency for a diagnosis involving the same body system. Although we focused on same-body-system-diagnosis admissions, our results are consistent with the findings of Schore (1997), who studied all Part A health services use.

Finally, there was no demonstration impact on mortality during the home health episode. Not surprisingly, in this elderly, sick group of patients, the mortality rate in the four months following home health admission was high, at nearly 10 percent. The estimated treatment-control difference, however, was only 0.1 percent.

The claims data analyses included numerous patient-, agency-, and area-level control variables (Table II.2). The coefficients for each model are presented in Appendix A. Not surprisingly, patient-level variables such as age, history of cancer, history of diabetes, and prior hospitalization were significant predictors of death and hospitalization. Interestingly, State (for example, Florida,

TABLE III.4

ESTIMATED IMPACTS ON PROBABILITY OF INSTITUTIONAL ADMISSION FOR SAME-BODY-SYSTEM DIAGNOSIS AND MORTALITY

Dependent Variable	Unadjusted Control Group Mean (Percentage)	Estimated Impact (P-value)[b]
Admission for Same-Body-System Diagnosis		
Hospital Admission During 120 Days	19.3	0.1 (0.67)
SNF Admission During 120 Days	5.4	-0.3 (0.90)
Home Health Care Admission During 120 Days[c]	5.7	0.0 (0.12)
Mortality		
Death During 120 Days	9.0	0.1 (0.83)

SOURCE: Medicare claims data from the National Claims History File.

[a]Episodes are weighted to give each agency equal representation. The "agency equal" weight for all episodes delivered by agency i is $w_i = \frac{1/n_i}{k/n}$ where n_i is the number of (episode-level) observations in agency i, k is the number of agencies, and n is the total number of observations for all agencies.

[b]The estimated impacts, obtained from logit models, are the treatment-control differences in the expected probability that the binary dependent variable equals one. These are computed as the average difference between two predicted probabilities for each observation, one treating the observation as belonging to the treatment group, the other treating it as belonging to the the control group. The t-statistics in parentheses are for the coefficient on the treatment status indicator in the logit model. If this coefficient is significantly different from zero, then the odds that the dependent variable equals one if an observation were in the treatment group are significantly different from the odds that the dependent variable equals one if that observation were in the control group. The t-statistics are based on standard errors inflated to account for clustering and weighting.

[c]Admission to a home health care agency different from the index agency.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.

Illinois) was not important for death and hospital admission but was significantly correlated with same-body-system-diagnosis admission for SNF and other Home Health Care Agency, even after controlling for area indicators of access such as urban/rural location, number of nursing home beds per 100 persons over age 65, and number of physicians per 10,000 persons. The effect of State in our models is consistent with findings by others showing widespread and unexplained geographic variation in home health care usage (Schore 1995; and Welch 1996).

C. EFFECTS OF ALTERNATIVE SAMPLE WEIGHTS

As described in Chapter 2, Section B.4., because the unit of randomization and intervention is the agency, we weighted observations to give agencies equal representation. However, weighting observations so that agencies are represented equally may accord small agencies undue influence in the analysis. To investigate the sensitivity of our results to the weighting approach, we also analyzed the data using sample weights that represented each agency proportional to its size, as measured by its share of demonstration episodes (Appendix B). Similar results from the two weighting methods would suggest strongly that small anomalous agencies did not exert undue influence, and that the results may be broadly interpreted for policy purposes. Conversely, while dissimilar results under the two sample weighting schemes do not necessarily indicate that the main results are incorrect, they suggest that further analysis is required to draw appropriate inferences about the likely effects of a national program.

1. QA Data, Observations Weighted to Represent Agencies Proportional to Their Size

The only two significant treatment-control differences among the 17 Improvement measures in the QA data with agencies weighted proportional to size (Table B.1, Appendix B) were in Improvement in Behavior Problem Frequency, with a difference of roughly 7 percent ($p=0.09$), and in Improvement in Light Meal Preparation with a difference of -6 percent ($p=0.07$). The differences

for these two outcomes in the analysis with agencies weighted equally were 7 percent (p=0.18) and -1 percent (p=0.65) respectively. Among the 17 Stabilization outcomes, only Stabilization in Surgical Wound Status was significant at the .10 level, with a treatment-control difference of -1 percent (p=0.09, Table B.2, Appendix B). The corresponding difference in the analysis with agencies weighted equally was also -1 percent (p=0.15). Finally, as shown in Table B.3, Appendix B, there were significant treatment-control differences in hospital emergency room visits (-3 percent, p=0.02), emergency visits to physician offices (-1 percent, p=0.06), and any emergency visit (-3 percent, p=0.02). The treatment-control differences for the same outcomes in the analysis with agencies weighted equally were -3 percent (p=0.20) for hospital ER visits, -1 percent (p=0.26) for physician office visits, and -4 percent (p=0.02) for any visits. Many of the point estimates of treatment-control differences in the Improvement outcomes reversed signs in the analysis with agencies weighted proportional to size, but for only one outcome (Transferring) was the initial difference statistically significant. Three of the Stabilization outcomes also showed a reversal in sign in the analysis with agencies weighted proportional to size.

Despite the differences for a few outcomes, we feel the results of the alternative analysis with weights to represent agencies proportional to size are generally consistent with our main analysis with weights to represent agencies equally. In neither analysis is there compelling evidence of demonstration impacts on symptoms and functioning. Unfortunately, neither set of analyses is easily interpreted, because of the measurement biases inherent in the QA data.

2. Claims Data, Observations Weighted to Represent Agencies Proportional to Their Size

In both the alternative analysis of the Claims data with agencies represented proportional to size (Table B.4, Appendix B), and the main analysis with agencies represented equally (Table III.4), none of the estimated impacts were statistically significant. Moreover, the point estimates of impacts for

all outcomes in both analyses were small in magnitude, with most less than 0.5 percent. The largest estimated impact was 1.2 percent for Home Health Care Admission in the alternative analysis. The results from both weighting schemes are thus consistent. Because the Claims data outcomes are captured in a consistent fashion for both treatment and control agencies, the lack of any significant differences in either weighting scheme indicates unambiguously that the demonstration had no impact on any of the Claims data outcomes.

IV. CONCLUSIONS

In this interim report, we find that the Home Health Per-Episode Prospective Payment Demonstration appears so far to have had little impact on a wide variety of patient outcome and health service utilization measures of home health quality of care. These measures include Improvement and Stabilization on a long list of health status indicators from demonstration QA data; Reported Use of Emergency Health Services from QA data; same-body-system-diagnosis admissions to a hospital, a SNF, or another home health agency from Claims data; and mortality from Medicare enrollment data.

Although these early data suggest no program impacts on patient outcomes from changes in agency behavior, it is possible that impacts will emerge later in the demonstration. In addition, our statistical power to detect small impacts is limited. Furthermore, the analysis of the QA data consisted only of unadjusted comparisons of treatment and control group means. More important, the QA data suffers from a systematic measurement bias between treatment and control agencies both in the length of time that patients remain under observation and in the frequency of home visits, a bias that severely limits our ability to interpret differences in treatment and control agency outcomes. Fortunately, while the same-body-system-admission and mortality outcomes in the Claims data may be less sensitive, and perhaps less relevant, indicators of home health quality of care than measures of functional status, they avoid the biased measurement problems of the QA data. Thus, the lack of consistent significant demonstration effects in the QA data, the lack of significant demonstration effects in the separate and independent Claims data, and the robustness of results in both data sets to the alternative agencies-proportional-to-size weighting scheme are all separate

pieces of evidence that, taken together, lead us to conclude that the demonstration has had no impact on home health quality of care.

The report by Cheh et al., "The Impact of Prospective Payment on Medicare Home Health Use: Promising Results for a Future Program" (1997), shows that the demonstration led to a significant reduction in number of home health visits, while the report by Schore, "The Impact of Prospective Payment on Medicare Service Use and Reimbursement During the First Demonstration Year" (1997), shows no significant shift towards increased use of other Medicare services. Taken together, the current report and the findings of Cheh et al. and Schore indicate that, so far, the demonstration has succeeded in lowering the home health care resources used per episode of care without evidence of harm to patients.

The final report on demonstration impacts on quality of care will include all three years of the outcomes described in this report, the results of policy-relevant subgroup analyses, and agency surveys on structural measures of quality of care. The final report will also present the patient survey data, which include measures of patient self-reported health and satisfaction collected in a uniform fashion across treatment and control agencies to allow an unbiased comparison of the two groups.

REFERENCES

Cheh, Valerie. Preliminary evaluation report on per-episode impact on home health costs. Princeton, NJ: Mathematica Policy Research, Inc., forthcoming.

Cheh, Valerie, Christopher Trenholm, Randall Brown, and Barbara Phillips. "Preliminary Report: The Impact of Prospective Payment on Medicare Home Health Use: Promising Results for a Future Program." Draft report submitted to the Health Care Financing Administration. Princeton, NJ: Mathematica Policy Research, Inc., November 1997.

Health Care Financing Review. *Medicare and Medicaid Statistical Supplement*, 1997. Baltimore, MD: U.S. Department of Health and Human Services, October 1997.

Health Care Financing Review. *Medicare and Medicaid Statistical Supplement*, 1996. Baltimore, MD: U.S. Department of Health and Human Services, October 1996.

Phillips, Barbara, and Rachel Thompson. "Transition Within a Turbulent System: An Analysis of the Initial Implementation of the Per-Episode Home Health Prospective Payment Demonstration." Draft report submitted to the Health Care Financing Administration. Princeton, NJ: Mathematica Policy Research, Inc., February 1997.

Phillips, Barbara R., Randall Brown, Valerie Cheh, Amy Klein, Jennifer Schore, Robert St. Peter, and Craig Thornton. "Evaluation of the Per-Episode Home Health Prospective Payment Demonstration Design." Report submitted to the Health Care Financing Administration. Princeton, NJ: Mathematica Policy Research, Inc., September 1995.

Prospective Payment Assessment Commission. *State Variation in the Resource Costs of Treating Aged Medicare Beneficiaries.* Washington, DC: ProPAC, June 19, 1996.

Prospective Payment Assessment Commission. *Report and Recommendations to Congress.* Washington, DC: ProPAC, March 1, 1996.

Schochet, Peter. "The Effects of Per-Visit Rate Setting for Home Health Care on Patients' Use of and Reimbursement for Medicare-Covered Services." Report submitted to the Health Care Financing Administration. Princeton, NJ: Mathematica Policy Research, Inc., May 1995.

Schore, Jennifer. "Regional Variation in the Use of Medicare Home Health Services." In *Persons with Disabilities: Issues in Health Care Financing and Service Delivery,* edited by Wiener et al. Washington, DC: The Brookings Institution, 1995.

Schore, Jennifer. "The Impact of Prospective Payment on Medicare Service Use and Reimbursement During the First Demonstration Year." Draft report submitted to the Health Care Financing Administration. Princeton, NJ: Mathematica Policy Research, Inc., December 1997.

Shaughnessy, Peter, Kathryn Crisler, and Robert Schlenker. *Medicare's OASIS: Standardized Outcome and Assessment Information Set for Home Health Care.* Denver, CO: Center for Health Policy Research, 1995.

Shaughnessy, Peter, Kathryn Crisler, Robert Schlenker, and Angela Arnold. "Outcome-based Quality Improvement in Home Care." *CARING Magazine*, February 1995.

Shaughnessy, Peter, Kathryn Crisler, Robert Schlenker, and Angela Arnold. "Outcomes Across the Care Continuum: Home Health Care." *Medical Care*, vol. 35, no. 11, supplement 1997.

Shaughnessy, Peter, Kathryn Crisler, Robert Schlenker, Angela Arnold, Andrew Kramer, Martha Powell, and David Hittle. "Measuring and Assuring the Quality of Home Health Care." *Health Care Financing Review*, vol. 16, no. 1, 1994.

Teplitsky, Sanford V., and Mary Ann Janson (eds.). *Home Health and Hospice Manual: Regulations and Guidelines.* Owings Mills, MD: National Health Publishing, 1985-1992.

Welch, H. Gilbert, David Wennberg, and W. Pete Welch. "The Use of Medicare Home Health Care Services." *New England Journal of Medicine,* vol. 335, no. 5, August 1, 1996.

APPENDIX A

COEFFICIENTS FOR MODELS PREDICTING DEATH, AND HOSPITAL ADMISSION, NURSING HOME ADMISSION AND HOME HEALTH READMISSION

TABLE A.1

ESTIMATED COEFFICIENTS FOR LOGIT MODELS PREDICTING DEATH AND HOSPITAL ADMISSION FOR THE SAME BODY SYSTEM DIAGNOSIS AS ADMISSION TO HOME HEALTH CARE DURING THE 120-DAY AT-RISK PERIOD

	Death (P-value)	Hospital Admission (P-value)
Intercept	-3.15*** (0.00)	-1.67*** (0.00)
Treatment Agency	0.02 (0.83)	-0.03 (0.67)
Age		
Younger than 65	-0.43*** (0.01)	-0.00 (0.99)
75 to 84	-0.08 (0.39)	0.06 (0.32)
85 or older	0.20** (0.04)	-0.09 (0.16)
Female	-0.50*** (0.00)	-0.09* (0.06)
White	0.24** (0.02)	0.15* (0.06)
Original Reason for Medicare Old Age	-0.04 (0.75)	-0.13 (0.11)
Medical Conditions		
Cancer	1.57*** (0.00)	0.33*** (0.00)
Diabetes	0.13 (0.12)	0.32*** (0.00)
Cerebrovascular accident (stroke)	-0.18 (0.06)*	-0.01 (0.89)
Decubiti stage 3 or 4	0.61*** (0.00)	0.05 (0.58)
Need for Complicated Wound Care[a]	-0.34*** (0.00)	-0.09 (0.29)
Functional Limitations[b]		
Bathing	0.16* (0.06)	0.04 (0.46)
Eating	0.64*** (0.00)	0.15** (0.03)
Dressing	0.11 (0.22)	0.02 (0.70)
Toileting	0.24** (0.04)	0.05 (0.51)
Transferring	0.23*** (0.00)	-0.13** (0.02)
Preadmission Location: Hospital	-0.08 (0.35)	-0.10* (0.07)

TABLE A.1 *(continued)*

	Death (P-value)	Hospital Admission (P-value)
Had Medicare for Less than Six Months	-0.27 (0.35)	0.19 (0.23)
Ratio of Mean Agency Visits to Mean for All Demonstration Agencies	-0.21 (0.21)	0.14 (0.24)
Hospital-Based Agency	-0.06 (0.48)	-0.01 (0.86)
For-Profit Agency	-0.08 (0.52)	-0.04 (0.71)
Chain Member	0.02 (0.78)	-0.02 (0.75)
Agency Provided Fewer than 30,000 Visits in Base Year	0.01 (0.90)	0.01 (0.90)
Agency Located in Urban Area	0.02 (0.86)	0.05 (0.73)
Agency State		
Florida	-0.17 (0.13)	-0.02 (0.73)
Illinois	-0.00 (0.98)	0.16 (0.17)
Massachusetts	-0.18 (0.23)	0.16 (0.21)
Texas	-0.05 (0.69)	0.00 (0.99)
County-Level Means		
Number of nursing home beds per 100 persons over age 65	-0.02 (0.50)	-0.02 (0.35)
Number of physicians per 10,000 persons	-0.01 (0.26)	-0.00 (0.79)
Hospital occupancy rate	0.32 (0.41)	-0.36 (0.45)
Medicare reimbursement per beneficiary (in thousands of dollars)	0.00 (0.72)	-0.00 (0.55)
Length of Hospital Stay During Two Weeks Before Home Health (Days)[†]	-0.00 (0.41)	0.01 (0.18)
Any Skilled Nursing Home Facility Stay During Two Weeks Before Home Heath	-0.35*** (0.00)	-0.32*** (0.00)
Total Medicare Part A Reimbursement During Six Months Before Home Health (in Thousands of Dollars)[‡]	0.00** (0.04)	0.01*** ((0.00)
Total Number of Home Health Visits During Six Months Before Home Health Admission	-0.00 (0.73)	0.00** (0.02)
Hospital Admission During Six Months Before Home Health Admission	0.27*** (0.01)	0.49*** (0.00)
Number of Episodes	**51,313**	**51,313**

TABLE A.1 *(continued)*

SOURCE: Medicare Standard Analytic Files 1993-1995.

NOTE: Estimates were obtained from weighted logit model, with observations weighted to give each agency equal representation.

[a]Patient has wound that requires soaking, irrigation, or debridement.

[b]Patient requires some human assistance with or does not participate in activity.

[c]If patient was not hospitalized within two weeks before home health, days are set to zero.

[d]Includes reimbursement for inpatient, skilled nursing facility, hospice, and nondemonstration home health paid under Medicare Parts A and B.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

TABLE A.2

ESTIMATED COEFFICIENTS FOR LOGIT MODELS PREDICTING SKILLED NURSING FACILITY ADMISSION AND OTHER HOME HEALTH AGENCY ADMISSION FOR THE SAME BODY SYSTEM DIAGNOSIS AS ADMISSION TO HOME HEALTH CARE DURING THE 120-DAY AT-RISK PERIOD

	Skilled Nursing Facility Admission (P-value)	Other Home Health Care Agency Admission (P-value)
Intercept	-3.50***	-3.81***
	(0.00)	(0.00)
Treatment Agency	-0.01	0.19
	(0.90)	(0.12)
Age		
Younger than 65	-0.31*	0.04
	(0.08)	(0.83)
75 to 84	0.32***	0.07
	(0.00)	(0.40)
85 or older	0.33***	0.08
	(0.01)	(0.44)
Female	0.02	0.17*
	(0.76)	(0.06)
White	0.31**	-0.12
	(0.03)	(0.38)
Original Reason for Medicare Old Age	-0.17	-0.06
	(0.19)	(0.66)
Medical Conditions		
Cancer	0.20**	0.14
	(0.03)	(0.32)
Diabetes	0.13*	0.19
	(0.10)	(0.05)
Cerebrovascular accident (stroke)	0.16	0.15
	(0.14)	(0.17)
Decubiti stage 3 or 4	0.33	0.07
	(0.13)	(0.65)
Need for Complicated Wound Care[a]	-0.10	0.00
	(0.56)	(0.98)
Functional Limitations[b]		
Bathing	0.29***	-0.00
	(0.00)	(0.98)
Eating	0.19**	-0.08
	(0.04)	(0.48)
Dressing	0.10	0.37***
	(0.31)	(0.00)
Toileting	0.01	-0.04
	(0.94)	(0.83)
Transferring	0.03	-0.15
	(0.70)	(0.30)
Preadmission Location: Hospital	-0.04	0.05
	(0.64)	(0.73)

TABLE A.2 *(continued)*

	Skilled Nursing Facility Admission (P-value)	Other Home Health Care Agency Admission (P-value)
Had Medicare for Less than Six Months	0.22 (0.44)	-0.18 (0.50)
Ratio of Mean Agency Visits to Mean for All Demonstration Agencies	0.23** (0.05)	0.54*** (0.01)
Hospital-Based Agency	0.02 (0.83)	-0.31 (0.14)
For-Profit Agency	-0.09 (0.43)	0.00 (0.99)
Chain Member	0.12* (0.06)	-0.05 (0.66)
Agency Provided Fewer than 30,000 Visits in Base Year	-0.01 (0.96)	0.22* (0.10)
Agency Located in Urban Area	-0.15 (0.29)	0.13 (0.50)
Agency State		
Florida	-0.35*** (0.00)	0.23 (0.25)
Illinois	-0.27* (0.08)	-0.62*** (0.00)
Massachusetts	0.09 (0.54)	-0.64*** (0.01)
Texas	-0.54*** (0.00)	-0.20 (0.29)
County-Level Means		
Number of nursing home beds per 100 persons over age 65	0.00 (0.90)	-0.04 (0.34)
Number of physicians per 10,000 persons	-0.00 (0.53)	0.01 (0.44)
Hospital occupancy rate	0.47 (0.26)	-1.42 (0.11)
Medicare reimbursement per beneficiary (in thousands of dollars)	-0.00 (0.31)	0.00 (0.22)
Length of Hospital Stay During Two Weeks Before Home Health (Days)[c]	-0.00 (0.84)	-0.00 (0.98)
Any Skilled Nursing Home Facility Stay During Two Weeks Before Home Heath	0.16 (0.14)	-0.23* (0.09)
Total Medicare Part A Reimbursement During Six Months Before Home Health (in Thousands of Dollars)[d]	0.00 (0.24)	0.01*** (0.01)
Total Number of Home Health Visits During Six Months Before Home Health Admission	0.00* (0.10)	0.01*** (0.00)
Hospital Admission During Six Months Before Home Health Admission	0.20* (0.08)	0.05 (0.67)
Number of Episodes	**51,313**	**51,313**

TABLE A.2 *(continued)*

SOURCE: Medicare Standard Analytic Files 1993-1995.

NOTE: Estimates were obtained from weighted logit model, with observations weighted to give each agency equal representation.

[a]Patient has wound that requires soaking, irrigation, or debridement.

[b]Patient requires some human assistance with or does not participate in activity.

[c]If patient was not hospitalized within two weeks before home health, days are set to zero.

[d]Includes reimbursement for inpatient, skilled nursing facility, hospice, and nondemonstration home health paid under Medicare Parts A and B.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

APPENDIX B

IMPACT ESTIMATES WEIGHTED PROPORTIONAL TO AGENCY SIZE

TABLE B.1

MEANS OF IMPROVEMENT OUTCOMES FROM QUALITY ASSURANCE DATA, WEIGHTED PROPORTIONAL TO AGENCY SIZE

Improvement in:	Treatment Group (Percentage)[a]	Control Group (Percentage)[a]	Treatment-Control Group Difference (Percentage)	P-value[b]
Pain	54.5	55.5	-1.0	0.69
Pressure Ulcer Count	85.4	88.7	-3.3	0.21
Most Problematic Pressure Ulcer	84.2	88.3	-4.1	0.15
Surgical Wound Status	81.8	80.3	1.5	0.67
Dyspnea	45.2	47.9	-2.7	0.44
Urinary Tract Infection	76.9	79.0	-2.1	0.43
Urinary Incontinence or Catheter Present	47.5	45.0	2.5	0.41
Confusion	42.1	38.6	3.5	0.30
Behavior Problem Frequency	65.2	58.5	6.7*	0.09
Grooming	50.3	53.3	-3.0	0.38
Bathing	50.2	53.9	-3.7	0.20
Toileting	53.9	55.7	-1.8	0.44
Transferring	46.6	48.6	-2.0	0.50
Ambulation/Mobility	31.1	32.9	-1.8	0.24
Light Meal Preparation	42.9	49.1	-6.2*	0.07
Housekeeping	40.1	43.6	-3.5	0.15
Management of Oral Medications	34.5	37.4	-2.9	0.32

SOURCE: Demonstration Quality Assurance Data, CHPR.

[a]In the Agencies Weighted Proportional to Size analysis, each agency receives representation equal to its relative size. Ordinarily this would be equivalent to conducting the analysis without sample weights, but because agencies had been in the demonstration for differing amounts of time, each observation is scaled by the relative time the agency had been contributing data. The "agency share" weight is given by $w_i^s = \bar{t} / t_i$, where t_i is the length of time agency i had been contributing data as of the sample cutoff date, and $\bar{t}$ is the average of the t_i's across agencies in the sample. The values of t_i and $\bar{t}$ are different between the QA sample and the Claims sample because of the different time frames and varying agency start-up dates for QA data collection.

[b]The p-values are from chi-square tests based on standard errors inflated to account for clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.

TABLE B.2

MEANS OF STABILIZATION OUTCOMES FROM QUALITY ASSURANCE DATA, WEIGHTED PROPORTIONAL TO AGENCY SIZE

Stabilization in:	Treatment Group (Percentage)[a]	Control Group (Percentage)[a]	Treatment-Control Group Difference (Percentage)	P-value[b]
Pain	83.2	82.4	0.8	0.41
Ulcer Count	99.1	99.0	0.1	0.61
Most Problematic Ulcer	99.1	99.0	0.1	0.57
Surgical Wound Status	97.4	98.1	-0.7*	0.09
Dyspnea	79.2	78.5	0.7	0.60
Urinary Tract Infection	98.1	97.9	0.2	0.29
Urinary Incontinence or Catheter Present	95.3	94.6	0.7	0.31
Confusion	85.9	85.8	0.1	0.94
Behavior Problem Frequency	93	91.6	1.4	0.11
Grooming	86.7	87.8	-1.1	0.25
Bathing	83.4	84.5	-1.1	0.27
Toileting	91.2	92.0	-0.8	0.41
Transferring	88.2	88.7	-0.5	0.61
Ambulation	90.0	90.1	-0.1	0.89
Light Meal Preparation	88.5	89.8	-1.3	0.11
Housekeeping	75.3	76.5	-1.2	0.33
Medication Management	87.4	88.1	-0.7	0.47

SOURCE: Demonstration Quality Assurance Data, CHPR.

[a]In the Agencies Weighted Proportional to Size analysis, each agency receives representation equal to its relative size. Ordinarily this would be equivalent to conducting the analysis without sample weights, but because agencies had been in the demonstration for differing amounts of time, each observation is scaled by the relative time the agency had been contributing data. The "agency share" weight is given by $w_i^s = \bar{t} / t_i$, where t_i is the length of time agency i had been contributing data as of the sample cutoff date, and $\bar{t}$ is the average of the t_i's across agencies in the sample. The values of t_i and $\bar{t}$ are different between the QA sample and the Claims sample because of the different time frames and varying agency start-up dates for QA data collection.

[b]The p-values are from chi-square tests based on standard errors inflated to account for clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.

TABLE B.3

REPORTED EMERGENCY HEALTH SERVICES USE

Reported Emergency Visits to:	Treatment Group (Percentage)[a]	Control Group (Percentage)[a]	Treatment-Control Group Difference (Percentage)	P-value[b]
Hospital Emergency Room	8.9	11.4	-2.5**	0.02
Hospital Outpatient Clinic	1.0	1.4	-0.4	0.13
Physician Office	2.0	2.7	-0.7*	0.06
Any of Above	11.5	14.6	-3.1**	0.02

SOURCE: Demonstration Quality Assurance Data, CHPR.

[a]In the Agencies Weighted Proportional to Size analysis, each agency receives representation equal to its relative size. Ordinarily this would be equivalent to conducting the analysis without sample weights, but because agencies had been in the demonstration for differing amounts of time, each observation is scaled by the relative time the agency had been contributing data. The "agency share" weight is given by $w_i' = \bar{t} / t_i$, where t_i is the length of time agency i had been contributing data as of the sample cutoff date, and $\bar{t}$ is the average of the t_i's across agencies in the sample. The values of t_i and $\bar{t}$ are different between the QA sample and the Claims sample because of the different time frames and varying agency start-up dates for QA data collection.

[b]The p-values are from chi-squared tests based on standard errors inflated to account for clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.

TABLE B.4

ESTIMATED IMPACTS ON ADMISSIONS FOR SAME BODY SYSTEM DIAGNOSIS AND MORTALITY, WEIGHTED PROPORTIONAL TO AGENCY SIZE

Dependent Variable	Control Group Mean (Percentage)[a]	Estimated Impact[a]	P-value[b]
Admission for Same Body System Diagnosis			
Hospital Admission During 120 Days	18.9	1.2	0.32
Skilled Nursing Facility Admission During 120 Days	5.2	-0.4	0.38
Home Health Care Admission During 120 Days	5.0	-0.2**	0.04
Mortality			
Death During 120 Days	9.7	-0.3	0.50

SOURCE: Medicare Claims data from the National Claims History File.

[a]In the Agencies Weighted by Size analysis, each agency receives representation equal to its relative size. Ordinarily this would be equivalent to conducting the analysis without sample weights, but because agencies had been in the demonstration for differing amounts of time, each observation is scaled by the relative time the agency had been contributing data. The "agency share" weight is given by $w_i^s = \bar{t} / t_i$, where t_i is the length of time agency I had been contributing data as of the sample cutoff date, and $\bar{t}$ is the average of the t_i's across agencies in the sample. The values of t_i and $\bar{t}$ are different between the QA sample and the Claims sample because of the different time frames and varying agency start-up dates for QA data collection.

[b]The p-values are from t-tests based on standard errors inflated to account for clustering and weighting.

**Significantly different from zero at the .05 level, two-tailed test.